STUDIES OF HOMOLOGOUS RECOMBINATION
BETWEEN PLASMID AND CHROMOSOMAL DNA

ADRIAN J. HARWOOD

Ph.D.
University of Edinburgh
1988
ABSTRACT

A scheme was devised in which the integration of a promoterless hprt/neo fusion gene into a chromosomal hprt gene would disrupt the chromosomal gene whilst expressing the neo gene from the chromosomal hprt promoter. It therefore should be possible to isolate these cells as they would be 6TG- G418-resistant.

The mouse metallothionien I promoter of pML2dMmtneo was shown to express the neo gene in both E. coli and the mouse L cell line, A9. Two independently derived human hprt cDNAs were partially sequenced and shown not to differ from the each other or the published sequence. Microinjection into A9 cells demonstrated that p4aA8, which contains one of the cDNAs, produced HAT-resistant colonies and generated the cell line M13 which contains a single integrated molecule of p4aA8.

The cell line A2-4, which is a mouse/human hybrid containing a single human hprt gene on the der(X;21) chromosome, and M13 were chosen as cells in which to attempt the targeted integration experiments. The rates of loss of hprt activity from the A2-4 and M13 cell lines were established to be 1 in 10^7 and 1 in 10^6 cells/generation, respectively. It was also shown that frequency of G418-resistant colonies after transfection with a promoterless neo gene was approximately 1x10^-6. This gave an expected background frequency of 1x10^-3 and 1x10^-1 for the targeted integration into A2-4 and M13 cell lines respectively.

The plasmid pAH, which expresses an hprt/neo fusion gene in both E. coli and A9 cells, was constructed. The promoter was removed to create pAH(M') and this was used in further gene targeting experiments.

After co-introduction of p4aA8 and pAH(M') into A9 cells G418-resistant colonies were obtained at a high frequency. Analysis demonstrated that these resulted from homologous recombination between the hprt sequences of both plasmids. In addition, a fragment of pAH(M') that contained homology to p4aA8 at only one end also produced G418-resistant colonies at a high frequency. It was also shown that two recombination events were involved in the generation of the final integrated plasmid.

Transfection of pAH(M') into A2-4 and M13 produced 6TG- G418-resistant colonies at an approximate frequency of 1x10^-7. Unfortunately, it was not possible to conclusively establish that these resulted from targeted integration into the hprt gene.
ACKNOWLEDGEMENTS

Firstly, I would like to thank Chris Bostock for his patient supervision, and help in moving to Pirbright, which enabled this project to be carried out. I am grateful to everyone at the MRC Mammalian Genome Unit (1984-1985) for getting me started, and everyone at the Institute of Animal Health (Pirbright) for helping me continue. I especially thank: Peter Fantes, for being my "other" supervisor and liaison with the University; Robin Allshire, Linda Dixon and John McCauley for help and advice with molecular biology; Chris Rowlands for the computing; and Stewart McKay, Sheila McLaughlin and Hannah Gostling for help with the tissue culture. I would also like to thank Beryl Hutchinson, Fiona Berry and Dawn O'Keefe for help in the presentation of this thesis. Lastly, I wish to thank Janet Savin for all support, help and enthusiasm.
TABLE OF CONTENTS

Title (i)
Abstract (ii)
Declaration (iii)
Acknowledgements (iv)
Contents (v)
Abbreviations (vi)

CHAPTER 1 INTRODUCTION

1.1 General 1
1.2 DNA mediated gene transfer 2
1.3 The introduction of DNA into the genome 4
1.3.1 Entry of DNA into the nucleus 4
1.3.2 Establishment of cell lines 7
1.4 Design of a plasmid for transfection into mammalian cells 9
1.4.1 Stability 9
1.4.1A Autonomously replicating vectors 10
1.4.1B Segregating extrachromosomal vectors 12
1.4.1C Integrating vectors 12
1.4.2 Selectable markers 13
1.4.3 Gene expression 15
1.5 Insertional mutagenesis 17
1.6 Homologous recombination 20
1.7 Mitotic recombination in yeast 22
1.8 Homologous recombination in mammalian cells 25
1.9 Targetted integration into mammalian chromosomes 29
1.10 Nonhomologous recombination 34
1.11 Aim of the project 36
<table>
<thead>
<tr>
<th>MATERIALS AND METHODS</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 General procedures</td>
<td>38</td>
</tr>
<tr>
<td>2.2 Growth, transformation and preparation of DNA from E. coli</td>
<td>38</td>
</tr>
<tr>
<td>2.2.1 E. coli strains</td>
<td>38</td>
</tr>
<tr>
<td>2.2.2 Culture media</td>
<td>39</td>
</tr>
<tr>
<td>2.2.3 Antibiotics/selective agents</td>
<td>40</td>
</tr>
<tr>
<td>2.2.4 Transformation of E. coli</td>
<td>40</td>
</tr>
<tr>
<td>2.2.5 Preparation of plasmid DNA from E. coli</td>
<td>42</td>
</tr>
<tr>
<td>2.2.6 Preparation of single stranded M13 phage</td>
<td>45</td>
</tr>
<tr>
<td>2.3 Growth, transformation and preparation of DNA from mammalian cells</td>
<td>46</td>
</tr>
<tr>
<td>2.3.1 Mammalian cell lines</td>
<td>46</td>
</tr>
<tr>
<td>2.3.2 Growth and maintenance of cell lines</td>
<td>46</td>
</tr>
<tr>
<td>2.3.3 Transfection by the calcium phosphate precipitation method</td>
<td>48</td>
</tr>
<tr>
<td>2.3.4 Microinjection</td>
<td>48</td>
</tr>
<tr>
<td>2.3.5 Preparation of DNA from mammalian cells</td>
<td>49</td>
</tr>
<tr>
<td>2.4 DNA analysis</td>
<td>50</td>
</tr>
<tr>
<td>2.4.1 Restriction endonuclease enzyme digestion</td>
<td>50</td>
</tr>
<tr>
<td>2.4.2 Agarose gel electrophoresis</td>
<td>51</td>
</tr>
<tr>
<td>2.4.3 Field inversion gel electrophoresis (FIGE)</td>
<td>52</td>
</tr>
<tr>
<td>2.4.4 Hybridisation of radioactive probes to DNA immobilised onto Biotrace-RP</td>
<td>53</td>
</tr>
<tr>
<td>2.4.5 Polyacrylamide gel electrophoresis</td>
<td>54</td>
</tr>
<tr>
<td>2.4.6 End-labelling of restriction enzyme fragments</td>
<td>56</td>
</tr>
<tr>
<td>2.4.7 Sequencing</td>
<td>57</td>
</tr>
<tr>
<td>2.5 DNA manipulation techniques</td>
<td>61</td>
</tr>
<tr>
<td>2.5.1 Isolation of DNA fragments</td>
<td>61</td>
</tr>
<tr>
<td>2.5.2 &quot;Blunt ending&quot; restriction enzyme fragments</td>
<td>62</td>
</tr>
<tr>
<td>2.5.3 Dephosphorylation of digested DNA</td>
<td>63</td>
</tr>
<tr>
<td>2.5.4 Ligation</td>
<td>63</td>
</tr>
<tr>
<td>2.5.5 Ligation of linkers</td>
<td>64</td>
</tr>
<tr>
<td>2.5.6 Colony/plaque hybridisation</td>
<td>65</td>
</tr>
<tr>
<td>2.5.7 Oligonucleotide site directed mutagenesis</td>
<td>66</td>
</tr>
<tr>
<td>2.6 Synthesis of oligonucleotides</td>
<td>68</td>
</tr>
<tr>
<td>2.7 Computing</td>
<td>69</td>
</tr>
</tbody>
</table>

(vi)
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>71</td>
</tr>
<tr>
<td>3.2</td>
<td>The plasmid pML2dMmtneo</td>
<td>73</td>
</tr>
<tr>
<td>3.3</td>
<td>A computer search for neo promoters</td>
<td>75</td>
</tr>
<tr>
<td>3.4</td>
<td>Transfection of pML2dMmtneo</td>
<td>78</td>
</tr>
<tr>
<td>3.5</td>
<td>Hprt sequence</td>
<td>82</td>
</tr>
<tr>
<td>3.6</td>
<td>Introduction of p4aA8 into A9 cells</td>
<td>84</td>
</tr>
<tr>
<td>3.7</td>
<td>Discussion</td>
<td>85</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>4.1</td>
<td>Introduction</td>
<td>89</td>
</tr>
<tr>
<td>4.2</td>
<td>The stability of hp rt gene expression in the cell lines A2-4, M13 and 4a5</td>
<td>89</td>
</tr>
<tr>
<td>4.3</td>
<td>Transfection of p(Bam)neo into A9 cells</td>
<td>91</td>
</tr>
<tr>
<td>4.4</td>
<td>Discussion</td>
<td>95</td>
</tr>
<tr>
<td>CHAPTER 5</td>
<td>CONSTRUCTION OF THE HPRT/NEO FUSION GENE</td>
<td>Page</td>
</tr>
<tr>
<td>-----------</td>
<td>----------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>5.1</td>
<td>Introduction</td>
<td>98</td>
</tr>
<tr>
<td>5.2</td>
<td>Principles of the design of the</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>hpRT/neo fusion gene</td>
<td></td>
</tr>
<tr>
<td>5.3</td>
<td>Computer modelling</td>
<td>101</td>
</tr>
<tr>
<td>5.4</td>
<td>Insertion of a linker into the neo gene</td>
<td>103</td>
</tr>
<tr>
<td>5.5</td>
<td>Creating the hpRT fragment</td>
<td>104</td>
</tr>
<tr>
<td>5.6</td>
<td>Inserting the hpRT fragment into pLink</td>
<td>107</td>
</tr>
<tr>
<td>5.7</td>
<td>Testing the kanamycin resistance of the</td>
<td>108</td>
</tr>
<tr>
<td></td>
<td>constructs</td>
<td></td>
</tr>
<tr>
<td>5.8</td>
<td>Discussion</td>
<td>111</td>
</tr>
<tr>
<td>CHAPTER 6</td>
<td>TARGETTED INTEGRATION OF pAH(M⁻) INTO THE HPRT GENE</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>-----------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>6.1</td>
<td>Introduction</td>
<td></td>
</tr>
<tr>
<td>6.2</td>
<td>Introduction of pAH(M⁻) into mammalian cells</td>
<td></td>
</tr>
<tr>
<td>6.3</td>
<td>Homologous recombination between pAH(M⁻) and p4aA8 following microinjection</td>
<td></td>
</tr>
<tr>
<td>6.4</td>
<td>Homologous recombination between pAH(M⁻) and p4aA8 following calcium phosphate precipitation</td>
<td></td>
</tr>
<tr>
<td>6.5</td>
<td>Homologous recombination into a chromosomal site</td>
<td></td>
</tr>
<tr>
<td>6.6</td>
<td>Discussion</td>
<td></td>
</tr>
</tbody>
</table>

Page

116
116
118
123
127
129

(x)
CHAPTER 7

DISCUSSION

REFERENCES

Page
138
143

(xi)
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>6TG</td>
<td>6-thioguanine</td>
</tr>
<tr>
<td>A</td>
<td>Adenine</td>
</tr>
<tr>
<td>Aph</td>
<td>Aminoglycoside 3’ phosphotransferase</td>
</tr>
<tr>
<td>ARS</td>
<td>Autonomously replicating sequence</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine 5’ triphosphate</td>
</tr>
<tr>
<td>bp</td>
<td>Base pairs</td>
</tr>
<tr>
<td>BFV</td>
<td>Bovine papillomavirus</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine serum albumin</td>
</tr>
<tr>
<td>BUdR</td>
<td>5-bromodeoxyuridine</td>
</tr>
<tr>
<td>C</td>
<td>Cytosine</td>
</tr>
<tr>
<td>ºC</td>
<td>Degrees Celsius</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complementary DNA</td>
</tr>
<tr>
<td>CHO</td>
<td>Chinese hamster ovary</td>
</tr>
<tr>
<td>CIAP</td>
<td>Calf intestinal alkaline phosphatase</td>
</tr>
<tr>
<td>DAPI</td>
<td>4’,6-diamino-2-phenylindol-dihydrochloride</td>
</tr>
<tr>
<td>dATP</td>
<td>Deoxyadenosine 5’ triphosphate</td>
</tr>
<tr>
<td>dCTP</td>
<td>Deoxycytosine 5’ triphosphate</td>
</tr>
<tr>
<td>ddATP</td>
<td>Dideoxyadenosine 5’ triphosphate</td>
</tr>
<tr>
<td>ddCTP</td>
<td>Dideoxycytosine 5’ triphosphate</td>
</tr>
<tr>
<td>ddGTP</td>
<td>Dideoxyguanosine 5’ triphosphate</td>
</tr>
<tr>
<td>ddTTP</td>
<td>Dideoxythymidine 5’ triphosphate</td>
</tr>
<tr>
<td>DEAE</td>
<td>Diethylaminoethyl</td>
</tr>
<tr>
<td>dGTP</td>
<td>Deoxyguanosine 5’ triphosphate</td>
</tr>
<tr>
<td>dhfr</td>
<td>Dihydrofolate reductase</td>
</tr>
<tr>
<td>dH2O</td>
<td>Distilled water</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>dTTP</td>
<td>Deoxythymidine 5’ triphosphate</td>
</tr>
<tr>
<td>EBV</td>
<td>Epstein-Barr virus</td>
</tr>
<tr>
<td>Term</td>
<td>Description</td>
</tr>
<tr>
<td>------</td>
<td>-------------</td>
</tr>
<tr>
<td>E. coli</td>
<td>Escherichia coli</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylene diaminetetra-acetic acid</td>
</tr>
<tr>
<td>FIGE</td>
<td>Field inversion gel electrophoresis</td>
</tr>
<tr>
<td>G</td>
<td>Guanine</td>
</tr>
<tr>
<td>g</td>
<td>Grammes</td>
</tr>
<tr>
<td>h</td>
<td>Hours</td>
</tr>
<tr>
<td>HAT</td>
<td>Hypoxanthine Aminopterin/Amethopterin and Thymidine</td>
</tr>
<tr>
<td>HEBs</td>
<td>HEPES buffered saline</td>
</tr>
<tr>
<td>HLA</td>
<td>Human leucocyte antigen</td>
</tr>
<tr>
<td>hph</td>
<td>Hygromycin-B-phosphotransferase</td>
</tr>
<tr>
<td>hprt</td>
<td>Hypoxanthine:guanine phsphoribosyl transferase</td>
</tr>
<tr>
<td>HSV</td>
<td>Herpes simplex virus</td>
</tr>
<tr>
<td>IMP</td>
<td>Inosine monophosphate</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl-beta-D-thio-galactopyranoside</td>
</tr>
<tr>
<td>LTR</td>
<td>Long terminal repeat</td>
</tr>
<tr>
<td>K</td>
<td>1,000</td>
</tr>
<tr>
<td>Kb</td>
<td>Kilobase pairs</td>
</tr>
<tr>
<td>M</td>
<td>Molar</td>
</tr>
<tr>
<td>MHC</td>
<td>Major histocompatibility complex</td>
</tr>
<tr>
<td>mg</td>
<td>Milligrammes</td>
</tr>
<tr>
<td>ml</td>
<td>Millilitres</td>
</tr>
<tr>
<td>min</td>
<td>Minutes</td>
</tr>
<tr>
<td>mM</td>
<td>Millimolar</td>
</tr>
<tr>
<td>Mmt</td>
<td>Mouse metallothionien I gene</td>
</tr>
<tr>
<td>M-MuLV</td>
<td>Moloney-murine leukaemia virus</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger RNA</td>
</tr>
<tr>
<td>MTX</td>
<td>Methotrexate</td>
</tr>
<tr>
<td>neo</td>
<td>Aph II</td>
</tr>
</tbody>
</table>
OD$_{260}$ Optical density at 260 nanometres
OD$_{550}$ Optical density at 550 nanometres
$^{32}$P Phosphorous, isotope 32
PBS Phosphate buffered saline
RNA Ribonucleic acid
RNase Ribonuclease
rpm Revolutions per minute
s Seconds
$^{35}$S Sulphur, isotope 35
S. cerevisiae Saccharomyces cerevisiae
SDS Sodium dodecyl sulphate
SSC Standard saline citrate
SV40 Simian virus 40
T Thymidine
TAE Tris, acetate, EDTA
TBE Tris, borate, EDTA
TE Tris-HCl, EDTA
TEMED N,N,N',N'-tetramethyll-1,2-diaminoethane
tk Thymidine kinase
Tris Tris (Hydroxymethyl) aminomethane
uCi microCuries
ug microgrammes
ul microlitres
UV Ultraviolet light
X-gal 5-bromo-4-chloroindolyl-beta-galactoside
xgprt Xanthine:guanine phosphoribosyl transferase
XMP Xanthosine monophosphate

(xiv)
CHAPTER 1

INTRODUCTION
1.1. General

The "gene" is not only a concept, but is a physical entity, which as DNA can be manipulated. By introducing genetically engineered DNA into cells, it is possible to change their characteristics. Within the area of mammalian systems, this technology has made a number of processes possible such as the use of mammalian cells to synthesise on the industrial scale substances of biological and medical interest, to create transgenic animals and in the future, it may be possible to use gene transfer as a means of treatment for genetic disease. It has also greatly aided the study of mammalian systems at the molecular level. This project is concerned with the problem of how to target the integration of exogenous DNA into a predetermined genomic site.

There are now a number of ways to introduce DNA into mammalian cells (Section 1.2). Processes such as the calcium phosphate precipitation method and microinjection (Section 1.3) can be used to introduce a DNA vector which will integrate into the genome (Section 1.4). In addition to creating cell lines which maintain the introduced DNA through long term culturing, integration offers the possibility of directly interacting with the host cell genome. A potential danger of uncontrolled integration, however, is that host genes may be disrupted by the integration of the exogenous DNA and so integration is potentially mutagenic (Section 1.5).

By using the cellular mechanisms of homologous recombination the site of integration may be controlled (Section 1.6). Homologous recombination has been observed in mitotically dividing yeast cells and has been used to target the integration of plasmid DNA to a predetermined site (Section 1.7). Similar homologous recombination processes have been observed in mitotically dividing mammalian cells.
(Section 1.8) and in a number of cases it has been possible to target the integration of introduced DNA (Section 1.9). The major problem in mammalian cells is that there is a high background of nonhomologous recombination (Section 1.10) which makes it difficult to identify those cells which have integrated the introduced DNA through homologous recombination. The aim of this project is to design a plasmid vector which can be used to select for those cells that have integrated it into a predetermined site (1.11).

1.2. DNA mediated gene transfer

In mammalian cells a number of methods of gene transfer have been developed, for example the formation of hybrid cells (Norum & Migeon, 1974); chromosome mediated gene transfer (McBride & Ozer, 1973); microcell fusion (Fournier & Ruddle, 1977), viral infection and DNA mediated gene transfer. As DNA mediated gene transfer introduces purified DNA into mammalian cells, it is one of the most useful methods of gene transfer from the point of view of the molecular biologist.

Szybalska and Szybalski (1962) demonstrated that mutant tissue culture cells, sensitive to aminopterin, became resistant to the drug at a low frequency following exposure to purified DNA from wildtype cells. The efficiency of DNA uptake (or transfection) can be greatly increased by using the calcium phosphate precipitation method (Graham and Van der Eb, 1972). Using this method to transfer herpes simplex virus (HSV) DNA into thymidine kinase deficient (tk⁻) mouse L cells and then growing the cells on medium that contained hypoxanthine, aminopterin and thymidine (HAT) it was possible to select those cells that expressed the thymidine kinase (tk) gene. This demonstrated that DNA from a foreign source could be used to complement a mammalian gene.
and enabled the fragment of HSV DNA that contained the tk gene to be isolated (Wigler et al, 1977; Maitland et al, 1977; Bacchetti & Graham, 1977; Pellicer et al, 1978). The demonstrations that non-selectable genes can be introduced by cotransfection with a selectable marker (Wigler et al, 1978), and that eukaryotic genes can be transfected as part of a bacterial plasmid (Colbere-Garapin et al, 1979; Graham et al, 1980), have opened up the technique for a wide range of uses.

There are now a number of methods of DNA mediated gene transfer. Modifications to the original calcium phosphate precipitation method have been suggested (Gorman, 1983; Luthman & Magnusson, 1983; Chen & Okayama, 1987), and the use of other materials, such as DEAE-Dextran (Sompayrac & Danna, 1981), Polybrene (Chaney et al, 1986) and Poly-L-Ornithine (Bond & Wold, 1987), have been investigated as ways to enhance transfection. Although increasing the initial uptake, with most of these modifications, the introduced DNA is only transiently expressed.

The process of electroporation, in which electric fields are used to form transient pores in the membranes of cells, has been developed as a method to facilitate DNA entry into the cell (Neumann et al, 1982). The general consistency of the results and the ability to transfect cells which normally are not amenable to calcium phosphate precipitation may mean that this method will eventually become the most effective method of DNA mediated gene transfer.

Microinjection offers a more direct method of introduction, enabling recombinant DNA to be investigated on the single cell basis, and can be used to introduce DNA directly into the nucleus (Capecchi, 1980). This has an advantage over other methods as the injected DNA avoids passage through the cytoplasm, resulting in a higher efficiency.
of transfer than other methods. As each cell must be injected individually, the technique is time consuming and so limited by the rate of injection. In addition to use with tissue culture cells, microinjection has been used to introduce DNA into fertilised oocytes and hence generate transgenic animals (Gordon et al, 1980; Simons et al, 1987).

In this project, both calcium phosphate precipitation and microinjection techniques were used to establish stable cell lines which had integrated plasmid DNA. In the next section these processes are considered in more detail.

1.3. The introduction of DNA into the genome

Transfection of mouse L cells with the HSV-tk gene demonstrated that a maximum of only 1 in $10^3$ to $10^4$ of the cells integrate DNA into the genome in a way that it can be expressed and stably retained during cell division (Wigler et al, 1978, 1979b). It is necessary therefore to select in order to identify those cells which have taken up the DNA.

What limits this process? Transfection can be considered as two processes: the passage of DNA to the nucleus, this is circumvented when injecting DNA directly into it, and the establishment of a stable cell line. In the case of DNA that cannot replicate extra-chromosomally, this later process involves a process of integration into the chromosomal DNA.

1.3.1. Entry of DNA into the nucleus

The first question to ask is why does the calcium phosphate precipitate method increase the transfection frequency?
There are a number of possibilities:

i) the precipitate protects the DNA from degradation either in the external medium or within the cell;

ii) it presents the cell with DNA in a form more suitable for uptake, eg. the precipitate attaches to the cell surface and is therefore more readily taken up by endocytosis;

iii) it stimulates cellular uptake of DNA.

Graham et al (1980) have shown that the physical appearance of a precipitate is related to the transfection frequency. The formation of a precipitate with the ability to transfect at a high frequency is determined by parameters such as the rate at which the components are added together and the DNA concentration. This suggests that during precipitate formation the DNA becomes arranged in a way that facilitates transfection and argues against the chemical stimulation of uptake, (iii) above. It would seem that the precipitate either protects the precipitate or presents a form which is more readily processed.

Incubation of DNA in the presence of medium with or without serum and analysis of the products by gel electrophoresis showed that there were nucleases present in the serum. If the DNA was formed into a calcium phosphate precipitate it was protected from these nucleases (Loyer et al,1982). This supports possibility (i) above. Protection may not be the only role of the precipitate. The difficulty experienced in transfecting suspension cells by the calcium phosphate precipitation method suggests that another role may be to bring the
DNA out of solution onto the surface of adherent cells and where it can be taken up by endocytosis.

Loyter et al (1982) took the study further by using the dye DAPI, which becomes fluorescent after binding to double stranded DNA, to follow DNA on entry into the cell. Staining showed that after transfection, DNA could be seen in the cytoplasm of most cells. By using the fluorescent stain chlorotetracycline, which specifically stains calcium complexes, calcium phosphate/DNA complexes were seen in the cytoplasm in most cells, but in the nucleus in only in 1-5%. This observation suggests that although DNA is readily taken up by cells, one of the major limiting steps of the transfection process is the transfer of DNA from the cytoplasm into the nucleus.

Bostock and Allshire (1986) showed that only after microinjection did the majority of BPV-1 DNA molecules, which had been linearised with BmnHI, recircularise and retain the BamHI site after entry into the nucleus. After transfection using the calcium phosphate precipitation method, circular DNAs were formed, however the majority of these could not be digested by BamHI. This suggests that the DNA was not completely protected by the precipitate from nuclease digestion so that the BamHI "sticky" ends were lost. It was not clear whether this digestion occurs in the external medium or on transfer through the cytoplasm. The possibility of degradation of DNA after entry into the cell is consistent with the notion that DNA enters the cell via endocytosis, which would lead to the precipitate entering the lysosome and hence possibly being degraded by lysosomal enzymes. Treating cells with chloroquin, which interferes with lysosomal enzyme activity, has been shown to enhance the uptake of polyoma DNA and reduce the degradation of linear DNA (Luthman & Magnusson, 1983). Cellular degradation of the transfected DNA therefore may be one
factor that limits the amount of DNA that reaches the nucleus. Damage to the input DNA also is likely to be one factor leading to the "rearranged" forms that are prevalent after transfection by the calcium phosphate precipitate method (Simons, 1985).

1.3.2. Establishment of cell lines

Although microinjection of DNA directly into the nucleus greatly increases the frequency of stable colonies, of the 80% of cells that express the introduced gene immediately after injection only 5-10% form stable colonies (C. Bostock, personal communication). Similarly, of the 1% of cells where the DNA has reached the nucleus following transfection, only 1-10% become stably transformed.

After transfection with a mixture of carrier DNA and the plasmid pTKX-1 that contained the HSV-tk gene, the number of HAT-resistant cells dropped to 2% of the original over a 20 day period (Scangos et al, 1981). Probing Southern blots with the plasmid DNA indicated that the DNA was covalently linked to high molecular weight DNA. The fact that the plasmid DNA could be isolated from unstable colonies after Hirt extraction suggested that this high molecular weight DNA was extrachromosomal. Microcell fusion and segregation analysis showed that in the stable colonies a particular chromosome was associated with the transfected tk gene in each case. This suggested that stable lines contained the transfected DNA integrated into the chromosomal DNA.

Similarly, if HSV-tk DNA was cotransfected with ØX174 and stable colonies isolated after HAT selection, analysis of the DNA from each colony shows that, in addition to the tk gene, the cell also contained ØX174 DNA; the two DNAs being covalently linked (Wigler et al, 1979a). If the transfected cell lines were selected for the loss of the tk
gene by growth on 5-bromodeoxyuridine (BUDR), loss of the tk gene was accompanied by the loss of the OX174 DNA (Perucho et al, 1980). In situ hybridisation has been used to confirm that the transfected DNA integrated into chromosomal DNA (Robins et al, 1981; Huttner et al, 1981).

From these results, it was concluded:

i) when DNA enters the nucleus it becomes covalently linked into high molecular weight extrachromosomal structures;

ii) in some cases these structures become integrated into the genome, producing a colony that is stable in long term culture.

These structures have been termed transgenomes (or pekalosomes).

The fact that colonies can divide for a number of generations before losing the transfected DNA, has lead to the suggestion that the transgenome structure may undergo limited replication.

After integration a transfected gene does not necessarily become expressed. In a number of cases the stability of integrated genes has been investigated (Pellicer et al, 1980; Gebara et al, 1987). The active genes were lost at a low frequency ( < 10^{-6}) by deletion from the genome, however in some cell lines gene expression was lost at a high frequency ( > 10^{-2}) without loss of the integrated gene. In a few of these cases this phenomenon was associated with methylation of cytosine bases, however in the majority of cases it was not clear what caused the inactivation. It is possible therefore that inactivation will occur to some genes before a stable cell line has been established from the founder cell.
To summarise, for stable expression of a gene, associated with DNA that does not autonomously replicate, it must enter the nucleus, integrate into the genomic DNA and remain in an active form. The processes of entry, integration and inactivation act to limit the establishment of stable colonies. The combined effect of these limitations is the major barrier to the establishment of stable colonies, and so varying the amount of DNA introduced to the cell only affects the transfection frequency when the amount of DNA is low (see Section 3.4).

1.4. Design of a plasmid for transfection into mammalian cells

DNA is most easily manipulated as a plasmid, it is therefore useful to have plasmids that can be used directly in mammalian cell systems. There are three basic requirements for any plasmid (or vector): stability, selectability and correct gene expression.

1.4.1. Stability

There are three types of vectors that can be used to increase the stability:

i) vectors that replicate extrachromosomally to high copy number, so that after division at least one copy of the plasmid will have segregated into each daughter cell;

ii) vectors that replicate extrachromosomally to low copy number, and contain sequences that segregate a copy of the plasmid into each daughter cell;

iii) non-replicating vectors which must integrate in order to segregate into each daughter cell.
1.4.1.A. Autonomously replicating vectors

No extrachromosomally replicating elements, equivalent to the plasmids of bacteria, have been discovered in mammalian cells, and so a number of different animal viruses are being investigated as the basis for autonomously replicating vectors.

The papovavirus, SV40, has a covalently closed double stranded DNA genome of approximately 5200 bp. It lytically infects monkey cells and undergoes limited replication in rodent cells. The binding of the virus protein, the large T antigen, to the "origin of replication" is sufficient for viral replication. The "origin of replication" is situated in the proximity of early gene promoter element regulating sequences. Plasmids that contain this sequence, such as pSV plasmids (Subramani & Southern, 1983), not only may express genes from the promoter but also replicate when large T antigen is present. A number of helper (or COS) cell lines have been constructed which contain the large-T antigen gene integrated into the genome (Gluzman, 1981; Boast et al, 1983; Rio et al, 1985), and plasmids containing the "origin of replication" may be replicated up to $10^5$ copies per cell in these cells (Boast et al, 1983). These plasmids are not stable however unless maintained by selection (Tsui et al, 1982).

An additional problem found with the replication of SV40 DNA containing plasmids is that there appears to be a cis-acting "poison" sequence in pBR322 which needs to be lost for replication to proceed. The effect of this loss appears to create random deletions affecting the ability of plasmids to replicate in either E. coli or mammalian cells. This "poison" sequence has been mapped to a region close to the bacterial origin of replication, which has been deleted from the pML series of plasmids (Lusky & Botchan, 1981).
The papillomavirus, bovine papillomavirus (BPV), which has a
cova~ntly closed double stranded genome of 7950 bp, has been shown to
infect and transform mouse cells, where it remains extrachromosomal.
Further study showed that the 69% subgenomic fragment, produced by
digestion with the restriction enzymes BamHI and HindIII, contains all
that is necessary for extrachromosomal replication (Sarver et al.,
1981). Plasmids that contain the 69% subgenomic fragment can be used
to synthesise up to 1 ug of foreign protein/10^6 cells/day (Hsiung et
al., 1984).

As with SV40, the combination of plasmid and virus appears to
affect the stability and integrity of the vector. With BPV, however
the interference due to the plasmid sequences appears to extend beyond
the presence of the "poison" sequence removed from pML plasmids
(Allshire & Bostock, 1986).

Another series of autonomously replicating vectors are based on
the herpesvirus, Epstein-Barr virus (EBV). A plasmid which contains
the EBV cis-acting element oriP and a single trans-acting factor,
encoded by the EBNA-1 gene, replicates in a variety of cells (Yates et
al., 1985). The major problem with this plasmid, however, may be that
it is lost in the absence of positive selection.

High copy number vectors therefore have been constructed which can
replicate in the nucleus of mammalian cells, by using elements derived
from mammalian viruses. There have been problems, however, with
stability in the absence of selection and the effects of
rearrangements when in combination with bacterial plasmid. In addition
to solving these problems, factors, such as the range of cell types
permissive for replication, must also be considered when constructing
these types of vector.
1.4.1.B. Segregating extrachromosomal vectors

The second strategy for stability has not progressed very far in the mammalian system, the challenge is to construct a vector to a similar design as the yeast linear vectors. In addition to an autonomously replicating element, these contain telomeric sequences to maintain the stability of the ends of the molecule and a centromeric sequence which interact with the mitotic apparatus to ensure correct segregation of the vectors into both daughter cells (Burke et al, 1987).

1.4.1.C. Integrating vectors

There have been two approaches to constructing integrating vectors. Firstly, plasmids, which do not contain sequences that induce replication, can only form a stable cell line by integration into the genome. Transfection with plasmid DNA, that contains a selectable marker, and growth for a number of weeks on the appropriate selective medium, will produce stable lines that contain integrated plasmid. Plasmid integration has the advantage over other methods, as it is possible to co-transfect unlinked DNA molecules into a cell (Wigler et al, 1979a), it is therefore not necessary to construct a single vector containing all of the DNA sequences to be introduced. This makes it possible to use transfection as a means to isolate DNA, as for example was the case with mouse aprt (Lowy et al, 1980).

The second approach is to use a vector based on retroviruses as these integrate into the genome as part of their infectious cycle. Retroviruses have a single stranded RNA genome which on infection is converted to covalently closed circular DNA. The circular DNA form integrates via a site present at the junction of 5′ and 3′ long terminal repeats (LTRs) using proteins carried in the virion and
encoded by the virus (Panganiban & Temin, 1984). Helper cell lines have been constructed which contain integrated retrovirus genomes that lack the site required for packaging into the virion, the psi site (Cone & Mulligan, 1984). These lines can be used to replicate and package defective virus or LTR containing plasmids, such as pZIPneo, which once packaged can be used to infect and integrate into the genome of a second cell line. Systems based on retroviruses therefore offer an effective means of introducing foreign DNA stably into a cell. These vectors are being considered as vehicles for gene therapy, ie. the replacement of a missing or defective gene responsible for an inherited disease.

A second consequence of integration is that it offers the opportunity to modify directly the genome at the sequence level. Such modifications would not only be of interest for academic study, but also opens up the possibility of gene replacement therapy, ie. the correction of defective genes responsible for an inherited disease.

1.4.2. Selectable Markers

As already mentioned, in order to identify those cells which have taken up DNA it is necessary to have a selectable marker. There are two forms of markers, morphological and biochemical.

A morphological marker is one where the appearance of the cell, or daughter cells, which have taken up the DNA is changed. This can be a difference such as the change in growth characteristics due to oncogenes, eg. H-ras (Shih & Weinberg, 1982) or infection with transforming viruses, such as BPV. Alternatively, a change in the cell surface molecules expressed after the transfection can be seen by using antibodies to the altered surface. For example, the human melanoma cell surface glycoprotein gp130 was identified by
Figure 1.1  Selectable genes of nucleotide biosynthesis

Enzymes that can be used as selectable markers are written in lower case.

Substrates are written in upper case.

Inhibitors are boxed.

N.B Xgpt pathway is boxed in broken line as it is not normally a mammalian pathway.

ABBREVIATIONS: PRPP, 5′-phosphoribosyl-1–pyrophosphate; PRA, 5′-phosphoribosylamine; IMP, inosine monophosphate; MPA, mycophilic acid, XMP, xanthosine monophosphate; GMP, guanosine monophosphate; AMP, adenosine monophosphate; CTP, cytosine triphosphate; UDP, uridine diphosphate; dUMP, deoxyuridine monophosphate, FH₄, tetrahydrofolic acid.
transfection and screening with a monoclonal antibody attached to erythrocytes, i.e. rosetting (Albino et al 1985). In addition, markers, such as the gene encoding for the luciferase protein responsible for bioluminescence in the firefly, Photinus pyralis (de Wet et al, 1987), can be used to produce a colorimetric change in the transfected cell.

A biochemical marker is one which enables the transfected cell to grow on medium toxic to the untransfected cell. It has the advantage over the morphological marker in that it removes the background of untransfected cells enabling novel cell lines to be established from single cell derived colonies.

The original biochemical markers used in transfection studies of mammalian cells were those genes which complemented mutations of the nucleotide biosynthetic pathways, such as thymidine kinase (tk; Wigler et al, 1977; Maitland et al, 1977; Bacchetti & Graham, 1977; Pellicer et al, 1978) and hypoxanthine phosphoribosyl transferase (hprt; Graf et al, 1979; Willecke et al, 1979; Peterson & McBride, 1980; Jolly et al, 1983; see Figure 1.1). By virtue of their inability to carry out specific reactions these mutants, are able to grow on medium containing BUdR and 6-thioguanine (6TG) respectively; substrates, which in the wildtype, are metabolised to give toxic products. It is therefore possible to back-select against cell lines which have taken up DNA that contains these genes.

The technique of DNA mediated gene transfer would be severely limited if it were dependent upon the generation of mutants, therefore a number of dominant markers have been developed. A mutant hamster dihydrofolate reductase (dhfr) gene has been isolated, which expresses a protein with reduced affinity for the inhibitor methotrexate, and so is able to confer resistance to higher concentrations than usual (Wigler, 1980). A set of dominant selectable markers that can be used
in mammalian cells have been isolated from bacteria. The bacterial protein, xanthine:guanine phosphoribosyl transferase (xgprt) is analogous to the hprt protein except that it converts xanthine to xanthosine monophosphate (XMP) rather than hypoxanthine to inosine monophosphate (IMP; see Figure 1.1). If expressed in a mammalian cell, it confers resistance to the block in the conversion of IMP to XMP when grown in medium containing mycophenolic acid (MPA; Mulligan & Berg, 1980, 1981).

The bacterial enzymes aminoglycoside 3'phosphotransferase (Aph) I and II allow growth in the presence of aminoglycoside antibiotics such as neomycin or kanamycin. The aminoglycoside antibiotic G418, or geneticin, is toxic to eukaryotic cells, and it has been shown that expression of the Aph II gene (generally referred to as neo) from either the HSV-tk promoter (Colbere-Garapiri et al, 1981) or the early SV40 promoter (Southern & Berg, 1982) confers resistance to G418. Similarly, a bacterial plasmid has been identified which encodes the gene hygromycin-B phosphotransferase (hph) and this allows growth in the presence of the aminocyclitol antibiotic hygromycin. This gene has been successfully employed as a selectable marker in mammalian cells (Gritz & Davis, 1983; Santerre et al, 1984).

The selectable markers hprt and neo are used in this project.

1.4.3. Gene Expression

Any vector must contain those sequences which promote correct gene expression, so that a selectable marker can function or proteins can be produced from other encoded genes. As the processes of transcription and translation are different in prokaryotic and eukaryotic cells, the eukaryotic sequences must be carried as an
expression cassette appropriate to the system to which the vector is applied.

Rather than list the possible expression cassettes that may be used, that of the plasmid p4aA8 which is used in this project, will be discussed. The plasmid p4aA8 was used to isolate the human cDNA hprt gene (Jolly et al., 1983) by cloning into the vector of Okayama & Berg (1983). This results in a plasmid where the cDNA is flanked by SV40 control sequences so that it can be expressed.

The SV40 early gene promoter is located 5' of the insertion site. This promoter utilises cellular transcription processes and contains the following elements.

i) Approximately 30 bp "upstream" of the point at which transcription is initiated there is a region showing consensus to the TATA (or Goldberg-Hogness) box. This is required to ensure that initiation of transcription from the correct position (Benoist & Chambon, 1981).

ii) Approximately 70-110 bp "upstream" of the point at which transcription is initiated, there are two perfect and one degenerate 21 bp repeats which have a core sequence of GGGGGCGGGGC. These are known to be the binding site for the cellular transcription factor Sp1 (Dynan & Tjian, 1983a,b). This region is required to maintain the correct level of transcription.

iii) Further "upstream" than both of these elements, there are two 72 bp repeats. These contain a sequence which is able to enhance transcription in a position and orientation independent manner (Khoury & Gruss, 1983).
These three elements combine to form a promoter which efficiently transcribes most genes in most cell types.

In addition to promoter sequences the 5' SV40 region also contains an intron. Although there is no a priori reason for this to be included, and many expression cassettes do not appear to contain introns, for example the expression cassette of pCGBPv9 (Matthias et al, 1983), it may be that the possession of wildtype splice donor and acceptor sites acts to limit the extent of cryptic splicing. To the 3' of the insertion site a second SV40 region provides the signal for polyadenylation.

Translation is less of a problem for expression in eukaryotic cells than in prokaryotic cells. Firstly, the regions required for efficient translation are generally cloned as part of the cDNA or within the body of the genomic DNA. Secondly, prokaryotic genes, such as neo, are translated in eukaryotic cells.

The other plasmid used in this study expresses the neo gene from the mouse metallothionein I (Mmt) promoter. This promoter has been used to express the rat growth hormone in transgenic mice when induced by heavy metals (Palmitier et al, 1982). It is also expressed at a basal level, and in this project this was found to be sufficient to confer G418 resistance on the mouse L cell, A9 (see Section 3.4).

1.5. Insertional Mutagenesis

As discussed above, integrating vectors are an effective way of establishing stable cell lines (or transgenic animals). They may also be used in the future to administer gene therapy. The integration of foreign DNA into the host genome however may induce mutations.
Several recessive mutations in transgenic mice have been identified which are due to the integration of a retrovirus. These either were experimentally induced, such as the strains MOV13 (Jaenisch et al, 1983) and ld (Woychik et al, 1985), or occurred naturally as is the case with dilute (Jenkins et al, 1981). The MOV13 integration has been investigated in detail. The mutation occurred as a result of microinjecting infectious Moloney-murine leukaemia virus (M-MuLV) into an embryo at mid-gestation. Litters produced by heterozygous parents for this insertion never produce homozygotes, as the embryos die during gestation (Jaenisch et al, 1983). It has been established that the provirus is inserted into the first intron of the alphal(I) collagen gene resulting in transcription initiation of only 1-5% the wildtype level in some tissues (Briendl et al, 1984). The fact that this, and many other mutants, are due to integration into the 5' region of a gene that would appear to be expressed in the embryo, has led to the suggestion that integration may preferentially occur at actively transcribing sites. This could be a result of these regions being more exposed to cellular enzymes and foreign DNAs.

This phenomenon of insertional mutagenesis is not confined to retroviral insertion. There are cases of transgenic mice where injection of plasmid DNA has apparently caused recessive mutations, such as the developmentally lethal mutations of strains HUGH/3 AND HUGH/4 (Wagner et al, 1983) and the strain Myk-103, where the integrated plasmid is never transmitted by males (Palmiter et al, 1984).

Dominant mutants may also result from integration into the genome. Although no experimentally induced dominant mutants have been identified, some tumours have been found to result from the insertion of promoters adjacent to cellular oncogenes. For example, some bursal
lymphomas of chickens show an unusual class of mRNAs that have virus sequences joined "upstream" of c-myc sequences. It appears in these cases that an integrated retrovirus in the lymphoma has a defective 5' LTR promoter. This is thought to allow the 3' LTR to express the genomic DNA sequences "downstream" from the retroviral insertion site, which in the case of some bursal lymphomas is a c-myc gene (Fung et al, 1981). This has been termed promoter insertion.

The danger is that if integration is used to administer gene therapy or to create transgenic animals, the effect of insertional inactivation or activation by promoter insertion, may negate the improvement gained by the introduction of the gene. On the other hand, the generation of mutants by integration of a known piece of DNA is a useful way to randomly generate mutants of previously undiscovered genes, as it "tags" the mutant DNA and so makes it easier to isolate the gene. This approach has been used with many types of organism, eg. prokaryotes (Calos & Milleri, 1980), yeast (Roeder et al, 1980), Drosophila (Bingham et al, 1981) and plants (Martin et al, 1985).

The problems posed by integration, as described above, could be overcome by the directed integration of a piece of DNA into a predetermined point. This process of targeted integration could be used to directly repair a defective gene, or alternatively, to direct the integration into an unimportant site, for example, the genes for coat colour in transgenic animals. It is possible to manipulate DNA with relative ease before introducing it into the cell, so the most straightforward strategy would be to design a DNA vector which would be able to undergo targeted integration by utilising cellular factors. The simplest vector would be one which contains sequences homologous to the required integration site and therefore is integrated through a process of homologous recombination.
As retrovirus integration involves a nonhomologous recombination mechanism between a site at the boundary of the two LTRs and a random site in the genome, this would be expected to predominate over homologous recombination between sequences carried within the retrovirus vector and a homologous genomic site. The best choice for a vector that will integrate through a process of homologous recombination, therefore, is one based on a nonreplicating plasmid.

1.6. Homologous Recombination

At the molecular level, the term homologous recombination is used in its broadest sense to describe the events that lead to the exchange of sequence between two nucleic acid molecules when such exchange is mediated through the presence of homologous sequences. It has been used to describe a variety of exchange processes as diverse as recombination during meiosis and a number of virus exchange mechanisms, including rearrangements of some RNA viruses (Kirkegaard & Baltimore, 1986). Homologous recombination between DNA molecules within a cell can be categorised as either a site-specific or a general process.

Site-specific events are mediated by specific homologous sequences and occur independently of the process of general recombination. They can be involved in developmentally programmed changes, such as Hin mediated inversion in the H2 locus of Salmonella which activates expression of the flagella antigen (Zieg & Simon, 1980), or mating type switching in yeast (Klar, 1980). It is also seen in more autonomous events such as the Lox/Cre system of P1 phage (Sternberg & Hamilton, 1981), Lambda integration through the att site (Kitts & Nash, 1987) and Chi sites (Smith, 1983) and the 2-micron plasmid rearrangement in S. cerevisiae (Broach et al, 1982).
Figure 1.2  The Aviemore model for the mechanism of homologous recombination

a) A nick in one DNA duplex leads to a free strand of DNA.
b) The free strand of DNA invades the homologous (recipient) DNA duplex, binding to its complementary strand and displacing a D-loop of DNA. The 3' end of the nicked strand acts as a primer for DNA synthesis, which displaces the strand ahead of it.
c) DNA synthesis further displaces the strand ahead of it which anneals to its complement in the recipient DNA duplex and in turn displaces more DNA from the recipient DNA duplex. The DNA of the D-loop is degraded. This enzymatically driven phase of the process results in the formation of asymmetrical heteroduplex DNA.
d) The 5' and 3' ends of the single strands become covalently joined to form the structure known as the Holliday junction. The Holliday junction can move along the DNA duplexes by diffusion (branch migration), producing symmetrical heteroduplex DNA.
e) Separate duplexes can be reformed by cleavage (resolution) of the Holliday junction in either of two planes.
f) Resolution in one plane (i) results in crossing-over whilst resolution in the other (ii) does not. In all cases DNA strands have been exchanged in the heteroduplex regions.

This model can explain the aberrant segregations of 8-spored fungi. If the DNA sequences differ within the asymmetrical heteroduplex DNA, then 5:3 segregation may occur. If the DNA sequences differ within the symmetrical heteroduplex DNA, then aberrant 4:4 occurs. The occurrence of 6:2 segregation is explained by the process of mismatch repair, asymmetrical heteroduplex DNA may be repaired to 4:4 or 6:2 and the symmetrical heteroduplex DNA repaired to 5:3 or 6:2.
General recombination mechanisms occur between any homologous sequences and, in addition to the meiotic recombination of eukaryotic organisms, the term is used to describe events in mitotically dividing eukaryotic cells and at least three bacterial systems (Clark et al., 1984). A mechanism of general recombination in mitotically dividing mammalian cells is required to target integration of plasmid DNA.

Before discussing possible mammalian mechanisms, it is useful to consider the basic recombination process as deduced from the study of fungi. The eight-spored fungi, such as Ascobolus immersus, have been useful tools to investigate meiotic homologous recombination. As each spore is derived from one of the eight DNA strands present on the pair of homologous chromosomes during meiosis, the pattern of spores enables the processes that occur during exchange to be observed (Rossignol & Paquette, 1979; Hamza et al., 1981). Two alleles (ie. markers) on either side and some distance from the cross-over point segregate to produce four spores of one type and four of the other (ie. 4:4). As the markers get within a short distance of the cross-over point aberrant segregation occurs. Aberrant segregations have three forms; 6:2, 5:3 and aberrant 4:4 (ie. 3:1:1:3). The apparent non-reciprocal transfer of information, as seen with 5:3 is termed gene conversion. The Aviemore model (Holliday, 1964; Meselson & Radding, 1974) offers an explanation of these processes at the molecular level. The model is described in detail in Figure 1.2.

The important features to consider of the Aviemore model are that in all situations heteroduplex DNA is generated, which in conjunction with mismatch repair, can account for the aberrant segregations noted above. Resolution of the Holliday junction (see Figure 1.2e) can occur in two planes, resolution in one results in reciprocal exchange of the flanking markers whilst in the other it does not. In all cases there
is exchange of the DNA strands that have become part of the heteroduplex DNA through involvement in the formation of the Holliday junction or branch migration. The process of homologous recombination may lead therefore to changes in the DNA sequences without any apparent exchange of flanking markers. In this discussion the term reciprocal exchange will describe the situation where the flanking markers are exchanged, and gene conversion will describe the situation where the flanking markers are not exchanged.

Enzymes have been identified which may mediate the processes proposed by the Aviemore model. The RecA protein is a major requirement for bacterial recombination. In vitro it has been demonstrated to catalyse the formation of D-loops and promote heteroduplex DNA formation (Shibata et al., 1979; DasGupta et al., 1980). A RecA-like protein, termed RecI, has been isolated from the fungi Ustilago maydis (Kmieć & Holloman, 1982), and RecA-like activity has been observed in extracts from RPMI 1788 B-lymphocytes (Hsieh et al., 1986). The RecBCD enzyme (Exonuclease V) cleaves one strand of DNA at Chi sites and has been thought to initiate recombination by creating a piece of single stranded DNA (Ponticelli et al., 1985). It has also been possible to use cruciform DNA to simulate the structure of a Holliday junction and use it to identify enzymes that may be involved in resolution. This has led to the discovery of T4 Exonuclease VII (Kemper et al., 1984) and a yeast protein (West & Körner, 1985) which are able to cleave cruciform DNA.

1.7. Mitotic recombination in yeast

In Saccharomyces cerevisiae, mitotic recombination has been shown to occur spontaneously and is enhanced by exposure to UV and ionising radiation (Fabre & Roman, 1977; Fasullo & Davis, 1987).
Jackson and Fink (1981) showed that homologous recombination between two loci located within a single integrated plasmid occurred in approximately 1 in $10^5$ cells per generation. This was three orders of magnitude lower than that in cells undergoing meiosis. In addition, they showed that unlike homologous recombination during meiosis, where reciprocal exchange and gene conversion occurred with equal frequency, the majority of the recombination events were gene conversion events. The Rad52 mutant showed a reduced level of homologous recombination and this was due to a reduction in level of gene conversion. Homologous recombination by reciprocal exchange remained unaffected by the Rad52 mutation. Rad52 mutants are also defective in both meiotic and irradiation induced recombination (Game et al, 1980). This argues that there are multiple mechanisms of homologous recombination in yeast some of which share common features with each other and with the DNA repair mechanisms induced after irradiation.

Hinnen et al (1978) demonstrated the use of the non-replicating plasmid, pYeLEU10, a plasmid containing the wildtype LEU gene, to convert leu yeast to LEU at a frequency of $1 \times 10^{-7}$. When the integration site was analysed in the majority of cases (30/42) the plasmid had integrated at the leu locus. This could be explained as a reciprocal exchange between homologous LEU sequences of the plasmid and the chromosomal site.

Linearisation of the plasmids, by cutting within the homologous region increased the integration frequency ten to a hundredfold (Orr-Weaver & Szostak, 1981, 1983). A plasmid which contained homologous sequences to two different regions of the genome, showed targeted integration to whichever region contained the cut. If the region of homology was cut with two restriction enzymes to generate a "gap", the gap was found to be repaired. "Gapped" plasmids containing
Figure 1.3 The double strand break repair model for the mechanism of homologous recombination

a) A double strand gap is made in one DNA duplex and exonucleases generate 3' single strands.

b) One single stranded end invades a homologous DNA duplex and displaces a D-loop.

c) DNA synthesis is primed from the annealed strand and further enlarges the D-loop until it anneals to its complementary 3' single strand.

d) DNA synthesis primed from the other 3' single strand completes the repair of the gap. Two Holliday junctions are formed which may both undergo branch migrations.

e) Resolution of each Holliday junction in the same plane (i) results in gene conversion, whilst resolution of each in a different plane (ii) results in reciprocal exchange.
an autonomously replicating sequence (ARS) could be found which had repaired the gap by gene conversion, after introduction into the cell. The integration of linear plasmids was found to be RAD52-dependent whereas integration of circular plasmids was not, suggesting that the presence of a double strand break, or gap, in a plasmid may have stimulated a different, higher frequency, process of homologous recombination than if the plasmid was circular.

Szostak et al (1983) proposed a model, now termed the double strand break repair model, which stated that homologous recombination may be initiated by double strand breaks, (for a detailed description see Figure 1.3). This model differs from the Aviemore model as follows:

i) initiation of recombination is due to a double strand break in one of the DNA duplexes and 3' single strand ends are produced by exonucleases;

ii) one of the broken DNA strands is repaired by invading the intact DNA duplex and using its complementary strand as a template, this displaces a D-loop which then acts as a template for repair of the other broken strand;

iii) after DNA synthesis and ligation two Holliday junctions are formed and if both junctions are resolved in the same plane then gene conversion occurs, if not then crossover occurs.
In addition to the stimulation of plasmid integration, double strand breaks appear to be involved in mating type switching of *S. cerevisiae* and are generated by the HO endonuclease. This process is also RAD52 sensitive (Malone & Esposito, 1980).

To summarise, there are multiple processes of homologous recombination in yeast. A plasmid can integrate through a process of homologous recombination at a low frequency. Linearising the plasmid within a region homologous to a chromosomal site increases the frequency of recombination. This occurs through a different process than that with the circular plasmids, and this higher frequency homologous recombination process has the capacity to repair missing regions of DNA.

1.8. Homologous recombination in mammalian cells

Homologous recombination has been implicated in a number of phenomena of mitotically dividing mammalian cells, such as the generation of MHC variants (Potter et al, 1987) and the generation of extracellular covalently closed circular DNA (Kiyama et al, 1987). In the past it has been difficult to observe chromosomal recombination in tissue culture cells due to the lack of suitable markers. Observations of Chinese hamster ovary (CHO) cells have shown recombination between two heterozygous markers on chromosome 2 at a frequency of approximately $1 \times 10^{-5}$ (Wasmuth & Vock Hall, 1984). Studies on the chromosomal structure of normal and malignant cells from patients suffering from retinoblastoma led to the suggestion that homologous recombination was responsible for the generation of homozygosity of the tumorigenic allele in some cases (Cavenee et al, 1983). The problem of lack of markers has been overcome by the use of DNA probes, so that it has been possible to use recombination to map the location
of the locus giving rise to the tumour rhabdomyosarcoma (Scrabble et al, 1987).

As an alternative to looking for cellular markers, studies have been carried out using introduced markers. For example, a single plasmid containing two segments of the HSV-tk gene, one with a 5' deletion and the other with a non-overlapping 3' deletion, has been stably integrated into the genome of mouse L cells. Intramolecular homologous recombination between these two sequences could result in cells which carried a complete expressible HSV-tk coding sequence and which would, therefore, be resistant to HAT. HAT-resistant colonies arose at a frequency of between $1 \times 10^{-4}$ to $10^{-6}$. Analysis of the DNA showed that 80% of clones restored the functional tk gene by gene conversion and 20% by reciprocal exchange (Letsou & Liskay, 1986). These observations demonstrate that intra-molecular homologous recombination does occur in mammalian cells, and appears to be similar to the events occurring in mitotic yeast cells (Jackson & Fink, 1981). Other groups have found similar results (Smith & Berg, 1984; Lin & Sternberg, 1984).

When plasmids are introduced into the nucleus they form concatamers in which they are arranged in a head-to-tail array. One mechanism by which these concatamers may form is via high frequency homologous recombination (Folger et al, 1982). Homologous recombination between two plasmids, each of which carry a copy of a selectable marker that has been inactivated by a non-overlapping deletion or insertion, has been shown to occur at a high frequency, so that up to 20% of the cells receiving DNA regenerated a wildtype gene (de Saint Vincent & Wahl, 1983; Shapira et al, 1983; Small & Scangos, 1983; Folger et al, 1985a). This frequency is many orders of magnitude higher than the frequencies obtained when DNA is part of the
chromosome. Folger et al (1985a) demonstrated that up to an hour could elapse between the separate injections of the two component plasmids without affecting the frequency of recombination. After this period the frequency dropped. This suggests that, after introduction into the cell, the DNA is rapidly converted to a form that it less likely to recombine.

An alternative way of investigating recombination uses mammalian viruses. By introducing mutations or by constructing SV40/pBR322 hybrids with overlapping virus sequences, it is possible to score recombination as the production of wildtype virus. As with plasmids, the frequency of recombination has been found to be very high (Wake & Wilson, 1980; Subramani & Berg, 1983).

Shapira et al (1983) noted that the frequency of recombination is roughly proportional to the distance between the mutations. A more detailed study has been carried out using hybrid SV40/pBR322 plasmids in which an overlap in the amount of SV40 sequence allows homologous recombination to release the wildtype virus (Rubnitz & Subramani, 1984). They showed that the frequency of homologous recombination in monkey cells reduced linearly as the length of the overlap dropped to 214 bp, at which point transfection of the SV40/pBR322 hybrids resulted in 0.25% of the plaques as infection with the wildtype. Below 214 bp there was a sharp reduction in recombination, however it was still detectable with only a length of 14 bp of homology. A similar relationship between length of homology and frequency of homologous recombination was found in EJ cells (Ayares et al, 1986).

As in the case of yeast, a double strand break in one of the plasmids results in an approximate ninefold stimulation in the frequency of recombination (Kucherlapati et al, 1984, Brenner et al, 1985; Wong & Capecchi, 1986). Gaps of up to 104 bp have been shown to
be repaired (Brenner et al, 1985). Recombination to generate a wildtype tk gene was found to occur even after the insertion of heterologous sequences into the region of homology. The absence of the heterologous sequence from the product of the recombination was suggested to result from exonuclease degradation of that region to generate a gap which was then repaired by homologous recombination (Brenner et al, 1986).

Folger et al (1985a) investigated the structure of plasmids rescued from mouse L cells selected on G418 after microinjection with two plasmids, which contained different neo gene point mutations. It was found that all wildtype genes formed by a gene conversion or a double reciprocal exchange event. No sequences were found to contain the double mutation. In this case gene conversion may have predominated as the point mutations were close enough to the site of recombination to be lost either due to enlargement of a double strand break or by conversion as a result of being present in the heteroduplex DNA.

The ability of the system to efficiently mismatch repair was demonstrated by microinjection of heteroduplex molecules generated by annealing strands from the two test plasmids together (Folger et al, 1985b). Cellular repair mechanisms have also been observed to remove single stranded from heteroduplex DNA (Ayares et al, 1987; Weiss & Wilson, 1987).

When both input DNA molecules contain double strand breaks a hundredfold stimulation of homologous recombination is seen over the circular form (Lin & Sternberg, 1984; Wake et al, 1985; Anderson & Eliason, 1986). This has led to a model that suggests that recombination may be mediated by the annealing of single stranded ends generated on both DNA molecules. These ends are produced from the
Figure 1.4 The single strand annealing model for the mechanism of homologous recombination

This figure describes the single strand annealing model proposed by Lin et al (1987), the principle is the same as the model described by Wake et al (1985).

a) Both DNA duplexes are broken by double strand breaks.

b) 3' exonuclease exposes homologous 5' single stranded DNA (shown as thickened lines).

c) The two duplexes anneal through the homologous regions of the single strands.

d) The 3' ends of the DNA act as primers for DNA synthesis, which proceeds to the region where the two duplexes are annealed (shown as dotted line).

e) The unannealed 5' strands are degraded and the 3' end of the synthesised DNA strand is ligated to the 5' end of the annealed DNA strand.
double strand breaks, either by a 3'-5' exonuclease (Lin et al, 1984, 1987) or by DNA unwinding (Wake et al, 1985). This is detailed in Figure 1.4. This model is supported by the observations that when two plasmids were cut in a nonhomologous region recombination may still occur between homologous sequences, and the highest frequencies are achieved when the cuts are equidistant each side of the region of recombination (Anderson & Eliason, 1986; Lin et al, 1987). If all recombination proceeded through a mechanism of single strand annealing, then both regions of DNA involved in the recombination must contain double strand breaks. It is possible that generation of these result in the lower frequencies of recombination observed for recombination involving circular plasmids or chromosomal sites.

To summarise, homologous recombination can be observed to occur in mammalian cells. In contrast to the low frequency of homologous recombination between sites within chromosomal DNA, coinjected DNA molecules recombine at a high frequency. A double strand break stimulates homologous recombination and in the process may repair gaps, as in yeast. An alternative process of single strand annealing has been proposed to account for some recombination events. By analogy with the situation in yeast, it may be that there are multiple pathways of recombination in mammalian cells which depend upon the structure of the DNA involved in the recombination and the different models describe some of these different pathways.

1.9. Targetted integration into mammalian chromosomes

As discussed in Section 1.3, plasmids integrate into the chromosomes of tissue culture cells at a frequency up to 1x10^-3. As a general rule, plasmids which contain homology to genomic sites do not appear to integrate into the chromosomes of either tissue culture
cells or transgenic animals by homologous recombination (Berg et al., 1983; Lacy et al., 1983). The first report that homologous recombination may occur after transfection was the demonstration that mouse L cells exhibited the serological determinants of the donor DNA haplotype after transfection with a truncated class I HLA gene (Goodenow et al., 1983). This was reported as being due to homologous recombination, but as no DNA analysis was carried out, it was not possible to confirm this.

Further investigations have followed two lines. Firstly, the introduction of deleted SV40 viruses into COS cells enables the generation of wildtype virus, presumably via a process of homologous recombination (Shaul et al., 1985; Jasin et al., 1985). Up to 25% of transfected cells produce wildtype virus. The best results were obtained when the input DNA had been linearised. These studies did not allow the changes which occur at the chromosomal site to be investigated, so the actual processes that result in the conversion to wildtype cannot be studied.

Alternatively, cell lines which contain defective genes have been generated. Supertransfection or microinjection of these lines with the same gene containing a non-overlapping defect, would be expected to result in the generation of a wildtype gene by a process of homologous recombination between input and chromosomal genes. Lin et al (1985) made ten cell lines containing a defective tk gene. In one line correction of the mutant gene occurred after calcium phosphate precipitation of the non-overlapping mutant by a process of homologous recombination. This occurred at a frequency of approximately $1 \times 10^{-6}$. Lin and co-workers consider this to be five orders of magnitude lower than would be expected for transfection of the wildtype gene, and hence a very low frequency event. If the transfection frequency of the
wildtype gene is considered to be nearer $1 \times 10^{-3}$ to $10^{-4}$, which would be a more likely transfection frequency, then homologous recombination would occur in 1/100 to 1/1000 cells. This is consistent with later reports. This frequency of homologous recombination was achieved with only 320 bp of homologous sequence between the integrating and target DNA. What is not clear is why it was not possible to target integration into the target sequences of the other nine cell lines. One possibility is that in some way these regions of DNA are inactivated, so that wildtype tk can be formed by homologous recombination, but it is not expressed. A similar low frequency ($1 \times 10^{-3}$ to $1 \times 10^{-4}$ of all transfectants) has been observed after similar experiments with the neo gene (Smith and Berg, 1984).

An increased efficiency of homologous recombination between an introduced plasmid and one integrated in the genome (both containing non-overlapping defects of the neo genes) of approximately 1 in 1000 cells injected, has been reported following microinjection of plasmid linearised in the neo sequence (Thomas et al, 1986a). In this study the one neo gene contained a point mutation which introduces a new restriction enzyme site whilst creating an amber mutation, the other contained a 284 bp deletion at the 3' end of the gene. Analysis showed that although in many cases restriction enzyme fragments characteristic of the wildtype neo gene had been created during the recombination event, in no case was the reciprocal double mutant gene found. It would appear that gene conversion rather than reciprocal exchange predominated in this system. This is a similar result to that found when the plasmids were coinjected (Folger et al, 1985a) and again may be an artefact due to the design of the experiment, as the input plasmid may have had an inherently high probability of gene conversion in the neo region.
A number of clones were obtained that apparently contained no wildtype genes. These were produced at a frequency five orders of magnitude higher than the spontaneous reversion rate for the integrated plasmid and only with injections involving sequence homologous to the neo gene. These have been analysed further to show that insertions of one or a few nucleotides have generated a new translational initiation codon a short distance from the amber mutation (Thomas & Capecchi, 1986). This mutation is thought to restore the neo activity by allowing re-initiation of translation to occur, and so nullifying the amber mutation. The insertions appear to be related to the formation of heteroduplex DNA during the process of recombination, and so this has been termed heteroduplex induced mutagenesis. The generation of novel sequence changes were also noted in sequences close to, but not within, the region where single stranded loops were removed from heteroduplex DNA constructed in vitro after having been transfected into the cell (Ayares et al, 1987).

Gene conversion may be disproportionally high in comparison with reciprocal exchange, as observed for intramolecular recombination in integrated plasmids in both yeast and mammalian cells (Jackson & Fink, 1981; Letsou & Liskay, 1986). In order to integrate a plasmid into the chromosome it must be designed so that the exchange of the flanking markers is selected for. This has been done in three cases.

By introducing a plasmid linearised in a region homologous to the beta-globin gene into either EJ or lymphoblastoid cells, it has been possible to isolate clones that had integrated the plasmid into this gene (Smithies et al, 1985). These events occurred at approximately 1 in 1000 transfectants and therefore made the isolation of the cells that had undergone targeted integration laborious. In this case, it was found that selection for integrating plasmid, by selecting for
expression of neo, only produced colonies which had integrated into the beta-globin gene in the lymphoblastoid cell line. As beta-globin is not expressed in the EJ cell line, it is possible that after integration into this area the neo gene was inactivated. This is consistent with the notion of inactivation of the wildtype tk formed by homologous recombination in the experiment described by Lin et al (1985).

Thomas and Cappecchi (1987) designed two vectors to integrate into the genomic hprt gene of mouse ES cells. By selecting different positions of the double strand break with respect to the regions of homology to the hprt gene, they have been able to construct vectors which either replace exon 8 of the hprt gene with the neo gene, or insert the neo gene into exon 7. In both cases those cells which have had the integration targeted into the hprt gene could be selected as 6TG- G418-resistant colonies. Doetschman et al (1987) also used the hprt gene to target integration. In this case, a 5' deletion mutant was the target gene. By introduction of a plasmid that contained the missing region plus a region with greater than 2.5 Kb of homology to the remaining mutant hprt sequence, it was possible to use HAT selection to isolate those cells that had targeted the integration of the plasmid.

In the first two cases described, a targeted integration could be detected in up to 1/1000 cells integrating DNA. In the third case, the frequency of targeted integration was approximately 14% of the frequency obtained with transfection of pSV2neo. As pSV2neo may only be expressed in a limited number of cells, the potential number cells integrating DNA into sites other than that of the hprt gene may be higher.
To summarise, it is possible to construct DNA vectors which will integrate into chromosomal sites. The high frequency of nonhomologous recombination and the possibility that gene conversion may occur at a higher frequency than reciprocal exchange, means that care must be taken in the design of a vector.

1.10. Nonhomologous recombination

From the previous discussion, in mammalian cells, plasmids integrate into the genome through a process of homologous recombination in about 1 in 1000 cells receiving DNA. This frequency of recombination is similar to that in yeast and perhaps some processes are common to both systems. The difference between yeast and mammalian cells, therefore, rather than being the ability to recombine through homologous sequences, would appear to be the high frequency of non-homologous recombination (illegitimate recombination) that occurs in mammalian cells.

Nonhomologous recombination has been implicated in many types of rearrangements in somatic mammalian cells, from the developmentally controlled immunoglobulin and T cell receptor rearrangements (Alt & Baltimore, 1982; Chien et al, 1987) to unregulated chromosomal translocation (Gerondakis et al, 1984). At the molecular level, it has been proposed to be involved in "simple" events, such as the generation of deletion mutants, and in more "complex" phenomena, such as amplification (Meuth, 1988). These nonhomologous recombination events appear to proceed through a mechanism where the DNA is first broken and then joined to a different piece of DNA. In this process nucleotides may be lost or gained at the junction between the two fragments.
Nonhomologous recombination of transfected DNA has been investigated by introducing into CV-1 cells SV40 DNA which had been cut with various restriction enzymes in the intron of the large T antigen gene. As the SV40 DNA must recircularise in order to be infectious, analysing the virus from individual plaques, has given a survey of the structures at the nonhomologous junctions between the ends of DNA molecules with predetermined sequences and structures (Roth et al, 1985; Roth & Wilson, 1986). Nearly all junctions (95%) occur within 15 nucleotides of the terminus of the introduced DNA. The majority of junctions could be accounted for either by ligation of blunt or single stranded ends, with or without loss of nucleotides, or, in the case of junctions between two ends with 5' protrusions, by a process of annealing through short homologies of a few nucleotides. It was suggested that "filling in" or complete removal of "sticky" ends to create blunt ends was rare. In addition there were a few junctions that involved addition of terminal nucleotides and a few where a more complex situations may have occurred. These more complex events may have been produced by unwinding of the ends and annealing of the single strand DNA to sequences that contain short regions of homology ("pseudohomology"). This mechanism may be similar to the single strand annealing model that has been proposed for homologous recombination.

As transfected DNA is known to become damaged during its passage to the nucleus, this may act to generate ends of DNA which undergo nonhomologous recombination. In addition enzymes such as topoisomerase I (Bullock et al, 1985), can generate breaks in DNA which may undergo nonhomologous recombination, and ligases have been identified in mammalian cells which may join the ends (Arrand et al, 1986).
1.11. Aim of the project

In this discussion, it has been shown that it is possible to use homologous sequences to target the integration of introduced DNA into a predetermined site in mammalian cells. Two factors interfere with this process. Firstly, there may be a bias towards gene conversion rather than the reciprocal exchange which is needed for a plasmid to integrate through homologous recombination. Secondly, the high frequency of nonhomologous recombination makes it difficult to identify those cells which have integrated a plasmid through homologous recombination.

One way of selecting for integration through reciprocal exchange is to have one marker flanking the homologous region of the integrating DNA and one flanking the homologous region of the target, so that after reciprocal exchange, the two markers are brought together on the same piece of DNA. A set of markers that can be found throughout the genome are promoters which are active for gene expression. By integrating a piece of DNA that contains a promoterless selectable gene so that it is in the correct orientation to an active promoter, it should become expressed and hence possible to identify such events on the appropriate selective medium. If a piece of DNA with homology to the coding region of the target gene is positioned "upstream" of a promoterless selectable gene, then homologous recombination should result in expression of the selectable gene and make it possible to isolate those cells which have undergone targeted integration.

In this project, an hprt/neo fusion gene was designed and constructed so that a region of hprt coding sequence was inserted in frame into the 5' region of the neo gene. This was used to test the feasibility of the scheme described above. An advantage in using the
Figure 1.5  The scheme to select for targeted integration of a plasmid into the hprt gene

This figure illustrates a single reciprocal exchange between hprt regions of a plasmid, that contains a promoterless hprt/neo gene, and an active chromosomal hprt gene. The integration would be expected to result in both the expression of the fusion gene, and hence G418 resistance, and loss of the hprt gene, and hence the switch from HAT to 6TG resistance.
hpert gene as the target, was that integration into the gene would disrupt its coding sequence and hence generate hpert\(^{-}\) cells which could be selected on 6TG. Cells that have targeted the integration of the fusion gene into the hpert gene were expected to be 6TG- G418-resistant (Figure 1.5).

These experiments differed from those of Smithies et al (1985), Thomas and Capecchi (1987) and Doetschman et al (1987) not only in design of the vector, but also in that a much shorter length of sequence homology was used. In the experiments listed, greater than 2.5 Kb of homologous sequence was used in each case. The fusion gene constructed in this project, used the cDNA as a source of hpert homologous sequence and only had 480 bp of homology with the cDNA. As the genomic hpert gene contains nine introns (Patel et al, 1986) there was a further reduction in the length of homology between each exon, 25 bp, 107 bp, 184 bp, 66 bp, 18 bp, 83 bp for exons 1,2,3,4,5,6, respectively. As homologous recombination can be detected between as little as 14 bp of homologous sequence (Section 1.8) and from the results of Lin et al (1985), it is of interest to see whether shorter stretches of homologous sequence can be used to target integration into a wildtype chromosomal gene.

Chapter 3 describes the initial characterisation of the plasmid DNA that was used to construct the hpert/neo fusion gene (in Chapter 5). In Chapter 4, the assessment of the frequencies at which hpert activity is spontaneously lost from the target cells, A2-4 and M13, and a promoterless neo gene becomes expressed after transfection is described. This gave an indication of the background that was expected in the targeting experiment described above. Chapter 6 describes the progress made towards establishing that this scheme can be used to target the integration into the hpert gene.
2.1. General procedures

All molecular biological techniques were carried out using Maniatis et al. (1982) as a guide. Chemical reagents were purchased from BDH, unless stated otherwise. Enzymes were purchased from Boehringer/BCL, unless stated otherwise. Radioactive isotopes were obtained from NEN/Dupont. Distilled water (dH$_2$O) was made using the Millipore RO system, whilst the dH$_2$O used in enzyme buffers was made using the Millipore Milli-Q system. Phenol was purchased ready distilled (Rathburn Chemicals Ltd.) and buffered with 0.1% 8-hydroxyquinoline and 0.2% beta-mercaptoethanol in dH$_2$O. TE consists of 10 mM tris, pH 7.4 and 1 mM EDTA.

2.2. Growth, transformation and preparation of DNA from E. coli cells

2.2.1. E. coli strains

Strain HB101: [G/S, F$, hsdR514 (R$^{+}$ M$^+$), recA13, proA2, lacY1, thi, str$^{R}$, endA]

This strain was used as a recipient for plasmid DNA.

Strain DH5: [F$, hsdR17 (R$^{+}$ M$^+$), recA1, thi-1, supE44, gyrA96 relA1]

This strain was also used as a recipient for plasmids DNA. It is derived from the strain DH1, but transforms significantly better (Hanahan, 1985).

Strain JM101: [pro, thi-1.supE,F$, traD36, proAB, lacI$^{q}$Z, M15]

This strain was used as the host for M13mp vector constructs.
2.2.2. Culture media

Strains HB101, DH5 and JM101 were grown in liquid LB-broth and on LB-agar.

LB-broth contains:

1% bacto tryptone (Oxoid)
0.5% yeast extract (Oxoid)
0.5% glucose
0.5% NaCl

pH adjusted to 7.2 with NaOH.

LB-agar contains LB-broth plus 1.5% agar (Difco).

Top-agar contains LB-broth plus 0.7% agar (Difco).

Before making competent cells all strains were grown in SOB-broth.

SOB-broth contains:

2% bacto tryptone (Oxoid)
0.5% yeast extract (Oxoid)
10 mM NaCl
2.5 mM KCl
10 mM MgCl₂
10 mM MgSO₄

Before transformation of the strain JM101 with M13 phage DNA the presence of the F-plasmid was selected by streaking cells onto a glucose minimal medium plate.

A glucose minimal medium plate contains:

M9 salts
1 mM MgSO₄
0.1 mM CaCl₂
1 mM thiamine HCl
0.2% glucose
1.5% Agar (Difco)
M9 salts contain:

- 50mM Na$_2$HPO$_4$
- 20mM KH$_2$PO$_4$
- 20mM NH$_4$Cl
- 8.5mM NaCl

2.2.3. Antibiotics/Selective agents

Ampicillin (Sigma) was made as a 1000x stock in sterile dH$_2$O, and used at 50 ug/ml.

Kanamycin (Sigma) was made as a stock of 25 mg/ml in sterile dH$_2$O, and unless stated used at 50 ug/ml.

5-bromo-4-chloroindolyl-beta-galactoside (XGAL) was made as a 2% stock in dimethylformamide.

Isopropyl-beta-D-thio-galactopyranoside (IPTG) was made as a 100 mM stock in sterile dH$_2$O.

2.2.4. Transformation of E. coli cells

The following method for the preparation of competent cells was described by Hanahan (1985), and routinely gave transformation frequencies of 10$^6$ to 10$^7$ colonies (or plaques)/ug. of DNA. Four to five bacterial colonies were picked from a LB-agar plate (or glucose minimal medium plate, in the case of JM101), and used to inoculate 100 ml of SOB medium. On reaching an OD$_{550}$ of 0.5, the culture was cooled on ice for 10 min and transferred to two 50 ml polypropylene tubes (Falcon). The cells were harvested by centrifugation at 1,500 rpm for 10 min, resuspended in 33 ml of precooled RFL and incubated on ice for 15 min (in the case of DH5 and JM101 cells) or for 2 h (in the case of HB101 cells). The cells were then collected by centrifugation,
resuspended in 12.5 ml of RF2 and put on ice for 15 min. The competent cells could either be used immediately or stored. To store the cells they were flash frozen in liquid nitrogen as 1ml aliquots and stored at -70°C.

Before transformation the cells were thawed on ice for 15 min. A 100 ul aliquot of cells was added to the 30 ng. (or less) of DNA, in a volume of less than 25 ul of TE pH7.5 (or 1x ligation mix). The tubes were put on ice for 30 min and then "heat shocked" at 42°C for 90 seconds. If plasmid DNA was used in the transformation, the mix was diluted to 1 ml in LB-broth, and incubated at 37°C for 1 h. Generally, 200 ul of the transformation mix was added to 3 ml of top-agar, spread on an LB-agar plate containing the appropriate antibiotics and incubated at 37°C overnight to form colonies.

If M13 phage DNA was used in the transformation, the whole mix was added to top-agar, containing 40 ul of XGAL stock, 40 ul of IPTG stock and 200 ul of a saturated fresh overnight culture of JM101 cells. This was then spread onto a LB-agar plate that contained no antibiotics, and incubated at 37°C overnight to form plaques. Vector containing insert DNA could be identified by the change in colour, from blue to clear, of the plaque.

RF1 contains:

100 mM KCl
50 mM MgCl₂
30 mM potassium acetate
10 mM CaCl₂
15% glycerol

Adjusted to pH5.8 and filter-sterilised.
RF2 contains:

- 10 mM MOPS
- 10 mM KCl
- 75 mM CaCl$_2$
- 15% glycerol

Adjusted pH 6.8 and autoclave.

2.2.5. Preparation of plasmid DNA from E. coli

Plasmid DNA was prepared from E. coli using a method based on the alkaline lysis method of Birnboim & Doly (1979). As the process was the same for whatever scale of preparation, the volumes are put in brackets, in the order: maxiprep, midiprep and miniprep.

LB-broth (500, 250, 5 ml), containing the appropriate antibiotic was inoculated with bacterial cells, and incubated at 37°C for approximately 18 h whilst shaking at 200 rpm. The cells were centrifuged from the LB-broth in 250 ml polypropylene tubes (Sorval) at 5K rpm and 4°C for 5 mins, or, in the case of minipreps in plastic universals (Sterilin) at 2,500 rpm and 4°C for 5 mins. The pellet was resuspended in solution I (18ml, 1ml, 180ul). The midiprep was transferred to a 50 ml polypropylene Oakridge tube (Sorval) and the miniprep transferred to a microfuge tube. Lysozyme (Sigma), dissolved to a concentration of 20 mg/ml in solution I, was added (2ml, 25ul, 20ul) to the cells. These were then left at room temperature for 5 min. Solution II was added (40ml, 3ml, 400ul) whilst mixing, and the tubes were put on ice for 5 min. Solution III (20ml, 2.25ml, 300ul) was added and the tubes left for a further 15 to 60 min on ice. The tubes were centrifuged at 10K rpm for 20 min, the supernatant was removed and precipitated with 0.6 volumes of isopropanol. The precipitate was spun down by centrifugation at 8K rpm and resuspended.
in 2M ammonium acetate (10ml, 500ul, 500ul). This was put on ice for 5 min and then centrifuged at 8K rpm for 5 min. This removed the proteins. The DNA was precipitated from the solution by addition of 0.6 volumes of isopropanol and centrifugation at 10K rpm. The pellet was resuspended in TE (10ml or 100ul).

At this stage midipreps and minipreps were found to be pure enough to digest with restriction enzymes and to use in ligations. To use midiprep DNA for microinjection, it was further purified by (10ml, 500ul, 500ul) digestion for 15 min with RNPse at 60°C, layered on 4 ml of 1M NaCl and centrifuged in a SW50.1 rotor at 40K rpm for 6 h. DNA was sedimented to the bottom of the tube, and was recovered by pouring off the NaCl solution and resuspending in 400 ul of TE. The DNA was precipitated by the addition of 0.6 volumes of 5M ammonium acetate and 2 volumes of absolute alcohol. The pellet was washed in 70% ethanol, dried and resuspended in 100 ul of TE.

The DNA solutions resulting from maxipreps were neutralised by addition of 1M tris base and made up to 14ml. 14g of CsCl and 1.4 ml of a 10mg/ml solution of ethidium bromide were added, and the density of the resulting solution was checked to be 1.55 g/ml (refractive index=1.3860) with a refractometer. The solution was transferred to a 30 ml polyallomer tube (Beckman). The air space was filled with liquid paraffin and the tube sealed. The tube was spun at 40K rpm for 36 h in a 70Ti rotor running in a Beckman L8-55M centrifuge. At the end of the run the rotor was stopped without using the brake. The supercoiled plasmid band was visualised using a 300nm UV lamp, and removed using a needle and syringe. The DNA was transferred to a 15 ml polypropylene tube (Sarstedt) and the ethidium bromide extracted by shaking with an equal volume of isopropanol saturated with CsCl and centrifugation at 2000 rpm. The DNA, contained in the upper aqueous layer, was removed.
and reextracted four or five times. The plasmid DNA was precipitated by the addition of exactly 3 volumes of 70% ethanol, left for 30 min at room temperature and then centrifuged at 2000 rpm for 15 min, at room temperature. The pellet was resuspended in 400 ul of TE, and reprecipitated by the addition of 0.6 volumes of 5M ammonium acetate and 2 volumes of absolute alcohol. The pellet was washed in 70% ethanol, dried and resuspended in 400 ul of TE. The concentration was measured by the optical density at 260nm (1 OD$_{260}$ unit corresponds to 50 ug/ml).

The only difference in the preparation of DNA of the RF form of M13 phage from that of plasmid DNA was that the initial culture was set up differently. The supernatant from a 5 ml culture of JM101 that had been transformed with M13 DNA was used to inoculate 100 ml of LB-broth that contained 1ml of a saturated fresh overnight culture of JM101 cells. This culture was then grown for 5 h and used to inoculate an 500 ml culture which also contained 1 ml of a fresh overnight culture of JM101 cells, and this was grown overnight, as for plasmid preparations.

Solution I contains:

50mM glucose
10mM EDTA
25mM tris-HCl, pH 8.0

Solution II:

1% SDS
0.2M NaOH

Solution III:

3M potassium acetate
2M acetic acid, pH4.8
2.2.6. Preparation of single stranded M13 phage

Single stranded M13 DNA was prepared for sequencing or oligonucleotide site directed mutagenesis. 100 ml of LB-broth was inoculated with 1 ml of a fresh overnight culture of JM101 cells, and dispensed as 1.5 ml aliquots into sterile universal containers. Fresh M13 plaques were picked and used to inoculate each tube. These were incubated for 4.5 to 5 h at 37°C whilst shaking at 200 rpm. The cultures were transferred to microcentrifuge tubes and the bacteria harvested by centrifugation at 10K rpm for 5 min. The supernatants were transferred to another microcentrifuge tube and recentrifuged to remove any bacteria inadvertently transferred from the first tube. The phage particles were precipitated by the addition of 20%PEG-6000/2.5M NaCl to each supernatant. These were left at room temperature for 30 min, then spun at 10K rpm for 10 min. The supernatants were completely removed with a drawn out Pasteur pipette. The pellets were resuspended in 100 ul of TE, and 50 ul of phenol was added. These were left at room temperature for 15 min, spun at 10K rpm for 3 min and the aqueous layers removed. Traces of phenol were removed by a chloroform extraction and the aqueous phases were left to precipitate overnight following addition of 10 ul of 3M Sodium acetate and 250 ul of absolute alcohol. The single stranded DNA was collected by centrifugation at 10K rpm and washed in 100 ul of 70% ethanol. The pellet was dried and resuspended in 20 ul of TE.
2.3. Growth, transformation and preparation of DNA from mammalian Cells

2.3.1. Mammalian cell lines

The A9 cell line is an hprt⁻, aprt⁻ mouse fibroblast line, derived from L cells by Littlefield (1964). The A2-4 cell line is a human-mouse hybrid cell generated by Sendai virus-induced fusion of an A9 cell with a human cell. This cell line contains the der X;21 chromosome which contains the short arm of chromosome 21 translocated onto Xp21-qter region (Worton et al, 1984). This does not interfere with the hprt gene which is situated at position Xq26-q27.3.

At the beginning of the project both lines were plated at low density and sub-cloned cell lines established. The same clone of A9 and A2-4 was used throughout the project.

2.3.2. Growth and maintenance of cell lines

Cells were grown as monolayer cultures in plastic tissue culture flasks and petri dishes (Falcon) at 37°C in an atmosphere of 95% air: 5% CO₂. All cells were grown on a alpha-modified minimal essential medium (Gibco) lacking nucleosides and deoxynucleosides, but supplemented with 10% foetal calf serum (Seralab), 100μg/ml penicillin and streptomycin (Gibco). A9 cell medium was supplemented with 0.36 μM 6-thioguanine (6-TG). A2-4 cell medium, and the medium for cell lines with integrated copies of the plasmid p4aA8, were supplemented with HAT. The final concentration of HAT is 0.36 μM hypoxanthine, 0.01 μM amethopterin and 0.2 μM thymidine. HAT and 6TG were also used at these concentrations when used as selective agents.

Cell lines were sub-cultured when confluent by pouring off the medium, washing the monolayer twice with trypsin solution and then leaving the cells covered with a small volume of trypsin solution.
until the monolayer had completely detached. A fourfold excess of complete medium was added to the flasks to inactivate the trypsin and the cells dispersed into it. The suspended cells could then be diluted accordingly and used to set up new cultures.

Stocks of the A9, A2-4, transfectant and microinjectant cell lines were stored in liquid nitrogen. Approximately $1 \times 10^6$ trypsinised cells suspended in medium and centrifuged at 1,000 rpm. The supernatant was discarded, the cell pellet resuspended in 1.5 ml of freezing medium and 0.5 ml was put into each freezing vial (Nunc). The vials were frozen overnight by putting the vials in a polystyrene container and placing at -70°C, and then transferred to liquid Nitrogen storage.

Trypsin solution contains:

- 0.5g trypsin
- 1.0g glucose
- 0.1g EDTA

Added to 50 ml of saline D concentrate, made up to 1 litre, and adjusted to pH 7.7 with HCl.

1 litre of saline D concentrate contains:

- 0.24g phenol red
- 160.0g NaCl
- 8.0g KCl
- 0.9g $\text{Na}_2\text{HPO}_4 \cdot \text{H}_2\text{O}$
- 0.6g $\text{KH}_2\text{PO}_4$

Freezing medium contains:

- 95% foetal calf serum
- 5% dimethyl sulphoxide.
2.3.3. Transfection by the calcium phosphate precipitation method

Cells were transfected with plasmid in the absence of carrier DNA using an adaptation of the method described by Graham et al (1980). A solution of 2xHEBS (50mM Hepes, 280mM NaCl, 1.5 mM Na₂HPO₄, 0.01mM EDTA, pH7.15) was added dropwise to an equal volume of 2xDNA-CaCl₂ solution (1mM Tris pH8.1, 0.1mM EDTA, 250mM CaCl₂, 20ug of plasmid DNA per ml) in a 15ml polystyrene tube (Falcon). The 2xDNA-CaCl₂ was mixed continually throughout this process by bubbling air through a Pasteur pipette into the bottom of the tube. The precipitate was left to form at room temperature for 30-40 min. An aliquot, equivalent to 0.15 ug of DNA/cm² of surface area of the area on which the cells were growing, was added to the medium covering the cells. These had been plated into fresh non-selective medium (ie. medium without 6TG or HAT) the day preceding the transfection. The cells were incubated for 24 h and the following day the medium, containing the precipitate, was discarded and fresh non-selective medium put onto the cells. Selective medium was put on the cells 24 h later. Cells transfected with the plasmid p4aA8 were selected on HAT medium. Cells transfected with the neo gene were selected on alpha-modified minimal essential medium, 10% foetal calf serum, penicillin and streptomycin supplemented with 800 ug/ml of G418 (Gibco). The medium was changed every 4 days. After 2-3 weeks on selective medium colonies had formed and were picked using a sterile cloning ring.

2.3.4. Microinjection

Microinjection can be used to introduce plasmid molecules directly into the nucleus of the cell. It is a more efficient way of introducing DNA into the nucleus, but is more time consuming than transfection by the calcium phosphate precipitation method.
The microinjection was carried out by myself or Dr. C. J. Bostock. The arrangement of the microscope and micromanipulator was as described by Simons (1985). Cells were plated at low density (1x10^3 cells/cm^2) into non-selective medium on 35mm petri dishes. Injections of cells were carried out as described by Simons (1985). The DNA was loaded into the needle at a concentration of 10 µg/ml in sterile PBS. The cells were injected within a few days of plating. The cells were injected in one or two lines across the dish. Each injection was estimated to introduce between 5 and 50 molecules of plasmid. The medium of each plate was changed immediately after injection. On the day following the injection, medium containing HAT or G418 was put onto the cells. Subsequently medium was changed every 4 days. After 2-3 weeks on selection colonies had formed. These were picked using a cloning ring and expanded into cell lines.

2.3.5. Preparation of DNA from mammalian cells

DNA was prepared using a method based on that of Gross-Bellard et al (1973). Cells were grown to confluence in 75 cm^2 flasks (approximately 1x10^7 cells). The monolayer was washed once with PBS and 5 ml of lysis buffer (10mM Tris-HCl pH8.0, 10 mM EDTA, 0.5% SDS containing 100 µg/ml Pronase, Calbiochem-Boehringer Corpn.) was added and left overnight at 37°C. The lysate was removed, extracted twice with one volume of phenol/chloroform (1:1) and then extracted once with one volume of chloroform. The aqueous phase was transferred to a small glass beaker and 1/10 volume of 3M sodium acetate added. Two volumes of absolute alcohol was layered on top of the solution, and the DNA wound out on a glass rod. The DNA was washed in 70% ethanol and allowed to dry. The DNA was resuspended in 500 µl of TE pH7.5 in a
microfuge tube. DNA prepared using this method was pure enough for digestion with restriction enzymes.

2.4. DNA analysis

2.4.1. Restriction endonuclease enzyme digestion

Restriction enzyme digests were carried out in sterile microfuge tubes. In each case, a core buffer was used, either high, medium and low salt, as described by Maniatis et al (1982). Digests were usually carried out in volumes of between 20 and 100 ul depending on the amount of DNA. The amount of enzyme used was equivalent to a theoretical threefold over digestion for a 3-4 h incubation at 37°C. At the end of each digest the enzyme was generally heat inactivated, but in those cases where the enzyme was refactory to heat inactivation the reaction was terminated by the addition of EDTA to a final concentration of 10mM. Double digests were carried out with both enzymes present if they had a common buffer. If not they were carried out sequentially with a DNA precipitation step between each digests. At the end of the digest, the DNA was precipitated by the addition of 0.6 volumes of 5M ammonium acetate and 2 volumes of absolute ethanol. The precipitation was cooled to -20°C for 1 h and then spun at 10,000 rpm for 20 min. The pellet was washed in 70% ethanol, dried and resuspended in an appropriate volume of TE or gel loading buffer.

- **High salt buffer**: 100mM NaCl, 50mM Tris (pH7.5), 10mM MgCl₂, 1mM dithiothreitol
- **Medium salt buffer**: 50mM NaCl, 10mM Tris (pH7.5), 10mM MgCl₂, 1mM dithiothreitol
- **Low salt buffer**: 10mM Tris (pH7.5), 10mM MgCl₂, 1mM dithiothreitol
2.4.2. Agarose gel electrophoresis

Agarose gels were used to size the fragments produced after restriction enzyme digestion. The fragments were visualised either by ethidium bromide staining and illumination with UV light, or by transferring the DNA from the gel to Biotrace-RP membrane and hybridisation with a sequence-specific probe.

These gels consisted of concentrations of agarose varying from 0.5 to 1.0%. Each gel was made by melting the appropriate concentration of agarose (Sigma, type I) 200 ml in 1xTAE (40mM Tris-acetate, 1mM EDTA, pH8.0) and contained 0.5 ug/ml ethidium bromide. Gels were set in a 25x20 cm gel tray (BRL), and run in a BRL H4 horizontal electrophoresis apparatus containing 1xTAE buffer. The samples were resuspended in 1xloading buffer (15% Ficoll, 0.5M EDTA, 0.25% orange G) and loaded into 6mm x 1mm slots. Gels were usually run at 0.5 volts/cm for approximately 18 h.

After electrophoresis the gels were placed on a Fotodyne transilluminator, illuminated with UV light (300nm) and photographed through a red filter (Hoya 25A) with a Polaroid MP4 Land camera, onto either Polaroid 667 or Kodak Technical Pan film. The Kodak Technical Pan film was developed by a Kodak X-OMAT M20 processor.

Mini-gels were run to check on the progress of restriction enzyme digests. These were 10cm x 7cm and used wells 3mm x 0.5mm. They ran in a Uniscience mini-gel box in 1xTBE (89mM tris-borate, 89mM boric acid, 2mM EDTA) running buffer. These gels were run at 5 volts/cm, and took one to two hours to run.

Low melting point (LMP) agarose gels were made with LMP Ultrapure agarose (BRL) and 1xTAE. These were 5cm x 7.5cm x 1cm, with a 3cm X 1mm slot. These were run in a GNA-100 gel box (Pharmacia) at 2.5 volts/cm for three to four hours.
300 ng per track of a HindIII digest of Lambda (cI857 Sam7) DNA (BRL) was loaded on agarose gel to provide marker bands of the following sizes: 23,130 bp, 9,416 bp, 6,557 bp, 4,361 bp, 2,322 bp, 2,027 bp and 564 bp.

A microdensitometer (Joyce-Loebel & Co. Ltd.) was used to scan the "copy number" gel (Figure 5.12), creating a trace on paper of the intensities of the bands from a photographic negative. The relative intensities were calculated by cutting out and weighing each peak.

2.4.3. Field inversion gel electrophoresis (FIGE)

A number of electrophoresis systems have been developed for resolving large DNA molecules (Schwartz et al, 1984; Carle et al, 1984; Southern et al, 1987). The FIGE system continually reverses the direction of the current, over a time period of seconds (Carle et al., 1986). This has the effect of separating fragments ranging in size from tens to hundreds of kilobases. The "window" of resolution appears to be determined by the overall length of the pulse time and the ratio of the forward to backward pulse times. By increasing the length of the pulse times during a run, ie. have a gradient, or "ramp", of pulse times, the gel appears to resolve over a range of sizes, with the smallest size, determined by the shortest pulse time length and the largest size, determined by the longest pulse time length. The ramp was set to build up over an interval of an hour and was then repeated through the course of a run.

A 1% agarose gel was made as for a conventional gel, except that the buffer was 0.5xTBE and no ethidium bromide added. The gel was run in 0.5xTBE in a modified BRL H4 horizontal electrophoresis apparatus. The buffer was continually circulated through a cooling coil to maintain the running temperature at 14°C. The gel field was run at 300
volts, giving a voltage gradient of 7.5 volts/cm. The direction of the current and length of the pulses were controlled via a switch box and a Sinclair Spectrum microcomputer with a programme "gel drive" written and supplied by D. Green, Edinburgh.

In the case of the gel in Figure 3.5b, the following conditions were used. The forward pulses increased from 1.5s to 7.5s, and the reverse pulse from 0.5s to 2.5s, during the course of the ramp. Electrophoresis was for 19 h, after which the gel was stained for 30 min in 0.5xTBE containing 0.5 µg/ml ethidium bromide. The gels were photographed and blotted as for conventional gels.

Markers were made by ligating the plasmid pBR328/TK/SAT, digested with EcoRI, at a concentration of 1µg/ul for 20 min at 14°C. The reaction was stopped by heat inactivation. The resulting oligomeric series increased with increments of 9.3 Kb.

2.4.4. Transfer of DNA from agarose gels onto Biotrace-RP

After electrophoresis DNA fragments could be transferred to a solid support and fragments, that contained sequence homology to DNA labelled with $^{32}$P (the "probe"), identified by hybridisation (Southern, 1975).

After electrophoresis, the gel was depurinated in 0.25M HCl for 15 min, which increased the efficiency of transfer of high molecular weight DNA by fragmenting the single stranded molecules. The gel was then washed three times in dH$_2$O and transferred to 0.4M NaOH. After 20 min, the gel was placed onto a blotting apparatus filled with 0.4M NaOH. Parafilm was placed at the edge of the gel to prevent the flow of liquid by-passing the gel and membrane. A piece of Biotrace membrane was soaked in 0.4M NaOH and placed onto the gel. This was followed by a piece of Whatman 3MM paper soaked in 0.4M NaOH, a dry
piece of 3MM paper and a stack of paper towels (Kim Dri). The towels were weighed down with a glass plate and transfer allowed to proceed for 18 h, after which the membrane was removed, rinsed in 4xSSC until neutral, and baked in a vacuum oven at 80°C for 2 h.

20xSSC contains:

- 3M NaCl
- 0.3M tri-sodium citrate

2.4.5. Hybridisation of radioactive probes to DNA immobilised on Biotrace-RP

The baked membrane was immersed in 4xSSC for a few min and transferred to a hybridisation tube. 15 ml of prehybridisation solution was added to the tube and the membrane incubated at 68°C for 18 h on a rotating rack. The prehybridisation solution was replaced by 10 ml of hybridisation solution containing heat denatured $^{32}$P-labelled probe. The membrane was incubated in the same manner for a further 18 h. The hybridisation solution was poured off and the membrane floated out of the tube with cold 4xSSC. The probe which had bound non-specifically was removed by sequential washes in 3xSSC (once), 2xSSC (twice) and 0.2xSSC (once). All washes contained 0.1% of sodium pyrophosphate and 0.1% of SDS, and were carried out at 67°C for 30 min. The filters were allowed to air dry for a few min, sealed in Saran wrap and autoradiographed whilst still damp.

Probes were made using the random priming method of Feinberg & Vogelstein (1983,1984). 100 ng of the DNA for the probe was denatured by boiling for 5 min. This was then cooled rapidly and added to 10 ul of OLB buffer and 2 ul of nuclease-free BSA (BCL). The mixture was made up to 47 ul, and 20 uCi of $^{32}$P-alpha-dATP and 2 units of the Klenow fragment of E. coli DNA polymerase I were added. The mix was
left at room temperature for 3 h. Unincorporated $^{32}$P-alpha-dATP was removed by passing the probe through a Sephadex G-50 column.

Autoradiography was carried out by placing the membrane in a X-ray cassette with X-ray film and an intensifying screen. The film was either Fuji-RX or Kodak X-OMAT AR5, which has a higher sensitivity. Fuji-RX film was preflashed before use. The cassette was placed at -70°C for times ranging from a day to two weeks. The combination of preflashing and -70°C increased the sensitivity of detection of the $^{32}$P tenfold (Laskey & Mills, 1977). The film was developed in a Kodak X-OMAT M20 processor.

When membranes were to be reused, previously hybridised probe was eluted from the membrane by washing in 0.4M NaOH for 30 min at 45°C. This was followed by a wash in a solution containing 0.1xSSC, 0.1% SDS and 0.2M Tris-HCl, pH7.5. The membrane could be then be hybridised to another probe.

**OLB contains:**

1M Hepes, pH 6.6  
0.25M Tris-HCl  
25mM MgCl$_2$  
40mM beta-mercaptoethanol  
0.1mM of dCTP, dGTP, dTTP  
1.35 mg/ml of hexadeoxyribonucleotides  
(Pharmacia)

**20xSET contains:**

3M NaCl  
0.4M Tris-HCl (pH 7.8)  
20mM EDTA
Prehybridisation solution contains:

- 5xSET
- 5xDenhardt's solution
- 0.1% SDS
- 0.1% sodium pyrophosphate
- 1 mg/ml of heat denatured Salmon sperm DNA.

Hybridisation solution is:

- prehybridisation solution,
- plus 10% dextran sulphate (Pharmacia).

50xDenhardt's solution contains:

- 1% Ficoll (Sigma)
- 1% BSA, Fraction V (Sigma)
- 1% polyvinylpyrrolidone (Sigma)

2.4.6. Polyacrylamide gel electrophoresis

Polyacrylamide gels were used for checking linker-kinasing reactions, sizing small fragments of DNA (less than 1000 bp) and isolating small DNA fragments.

Acrylamide:bis-acrylamide solution was diluted in 1xTBE from a 30% stock solution to make a 50 ml solution of the required percentage. 375 ul of a 10% ammonium persulphate solution and 50 ul of TEMED (NNN’N’ tetramethyl-1,2-diaminoethane; Eastman-Kodak) was added, and the solution was poured immediately into a vertical gel apparatus (2 glass plates held 1mm apart by plasticard spacers and sealed with 1% agarose on three sides). A comb (teeth 5mm wide) made from 1mm thick plasticard, was inserted into the top edge. The gel usually set within 15 min of pouring.
The sample was loaded in 1xTBE that contained 2% Ficoll and 0.5% bromophenol blue and xylene cyanol. The rate of migration of these dyes depends on the percentage of the dye. Gels were run in 1xTBE for 2 to 3 h at 200 volts.

Polyacrylamide gels were were either stained and viewed on a transilluminator or dried down onto Whatman 3MM paper and autoradiographed.

30% acrylamide stock solution contains:

- 29g acrylamide
- 1g N’N’ methylene bis-acrylamide

in 100 ml of dH₂O.

2.4.7. **End-labelling of restriction enzyme fragments**

By labelling DNA with $^{32}$P and running it on a polyacrylamide gel it is possible to see fragments that are either too small or too low concentration to see with ethidium bromide staining. This labelling can be achieved by filling in 3’ recessed ends of fragments produced by digestion with certain restriction enzymes, using the Klenow fragment. This enzyme carries the 5’ to 3’ polymerase activity of DNA polymerase I. Alternatively, 5’ recessed or blunt ends can be labelled using T4 polynucleotide kinase.

To end label a digest with the Klenow fragment, at the end of a restriction enzyme digest, 1 unit of Klenow fragment, a mixture of 0.5 mM nucleotide triphosphates and 5 uCi of $^{32}$P-alpha-dATP were added and the mix left at room temperature for 30 min. The DNA was then precipitated with 0.6 volumes of 5M ammonium acetate and 2 volumes of absolute alcohol. The pellet was washed in 70% ethanol, dried and resuspended in TE.
T4 polynucleotide kinase catalyses the transfer of the gamma phosphate from ATP to the 5' termini of DNA molecules. The DNA was incubated for 15 min with 20 uCi of ^3P-gamma-ATP, 10 units of T4 kinase and one tenth volume of 10x kinase buffer (0.7M Tris-HCl, pH7.6, 0.1M MgCl₂, 50mM dithiothreitol). 10 ul of 10 units of T4 polynucleotide kinase and 1mM ATP in 1x kinase buffer was added and the mix incubated for a further 15 min at 37°C. The DNA was then precipitated with 0.6 volumes of 5M ammonium acetate and 2 volumes of absolute alcohol. The pellet was washed in 70% ethanol, dried and resuspended in TE. In the case of linkers, the labelled linkers were used directly in a ligation. To label restriction enzyme fragments it was necessary to first dephosphorylate the fragment (see Section 2.5.3).

2.4.8. Sequencing

The dideoxy chain termination method (Sanger et al, 1977) was used to sequence hprt cDNA subcloned into the M13mp8/19 vectors (Messing et al, 1977). The protocol used was based on the procedures for carrying out the sequence reactions in microtitre plates as described by Bankier et al (1988).

Four wells of a disposable 96 U-well microtitre plate (Gallenkamp) were designated "A", "C", "G", "T", for each template. To each well 2 ul of primer/TM mix was added. This contains 1ul of M13 universal primer (0.2 pmol; Boehringer/BCL), 1 ul of TM and 7 ul dH₂O for each clone. 2 ul of each template was added to each set of 4 wells, "A" to "T". The plate was covered with plastic film, a brief centrifugation used to mix the contents and the plate incubated for 2 h at 55°C. This allowed the primer to anneal to the template.
Into each well 2 ul, of the appropriate nucleotide mix was dispensed. A mix of 0.4 ul (4 uCi) of $^{35}$S-alpha-dATP, 1 ul of 100mM dithiothreitol, 0.2 ul of the Klenow fragment (5u/ul) and 5.6 ul dH$_2$O for each template was made, and 2 ul dispensed into each well. The plate was covered, the contents mixed by brief centrifugation and the plate incubated for 20 min at 37°C. To each well, 2 ul of Chase solution was added, mixed and the incubation continued for a further 15 min. At this point the reaction mixes could be stored at -20°C before running on a gel. To run on a gel, 4 ul of formamide dye was dispensed into each well and the plate was placed uncovered into an oven at 90°C for 30 min. After the samples had concentrated to a volume of approximately 2 ul, the entire contents of each well were run on a sequencing gel.

The sequencing gels were 6% polyacrylamide gels, made from a stock of 28% acrylamide and 2% bis-acrylamide, that contained 7M urea. These were made as described in Section 2.4.6, and were 0.4 mm thick. The gel was assembled into a BRL sequencing gel apparatus with a thermostatic plate. The samples were loaded between the teeth of a 5mm "sharks tooth" comb, the 4 bases of each template grouped together. The gels were run in 1xTBE at a constant 60 Watts. Gels were run between 2 and 4 h depending on the distance that the sequence of interest was from the priming site. At the end of the run the gel was fixed in 10% methanol/10% acetic acid for 15 min. The gel was then transferred to Whatman 3MM paper and dried. The dried gel was auto-radiographed with X-OMAT AR5 film for one to three days. The sequence was then read by eye.

**TM contains:**

100mM Tris pH8.5
50mM MgCl$_2$
Nucleotide mixes

"A" mix contains:
500 ul 0.5 mM dTTP
500 ul 0.5 mM dCTP
500 ul 0.5 mM dGTP
1 ul 10.0 mM ddATP
500 ul TE

"C" mix contains:
500 ul 0.5 mM dTTP
25 ul 0.5 mM dCTP
500 ul 0.5 mM dGTP
8 ul 10.0 mM ddCTP
1000 ul TE

"G" mix contains:
500 ul 0.5 mM dTTP
500 ul 0.5 mM dCTP
25 ul 0.5 mM dGTP
16 ul 10.0 mM ddGTP
1000 ul TE

"T" mix contains:
25 ul 0.5 mM dTTP
500 ul 0.5 mM dCTP
500 ul 0.5 mM dGTP
50 ul 10.0 mM ddTTP
1000 ul TE

Chase solution contains:
0.5 mM of each dNTP
Formamide dye contains:

100 ml formamide (deionised with mixed bed resin)
0.1g xylene cyanol F.F.
0.1g bromophenol blue
2 ml 0.5 M EDTA. pH8.0

2.5. DNA manipulation techniques

2.5.1. Isolation of DNA fragments

Two methods were used to isolate DNA fragments from gels depending on their size. Large fragments were isolated by electrophoresis in low melting point agarose as a gel slice that contained the DNA fragment. Small fragments were eluted from a slice cut from a polyacrylamide gel.

To isolate a fragment greater than 1 Kb, up to 50 ug of digested DNA was run on a low melting point agarose gel. After running the gel for 3-4 h, the band was visualised on a long wavelength UV (300nm) transilluminator, cut out with a sterile scalpel blade and put into a sterile microfuge tube. The DNA and agarose was diluted with an equal volume of TE and melted at 68°C. The solution was cooled to 37°C and 50 units of agarase added. Agarase (Calbiochem-Boerhinger Corpn.) is purified from Pseudomonas atlantica, and stored at 5 units/ul in a 50mM potassium phosphate buffer (pH6.9). The solution was incubated for 2 h at 37°C. At the end of this period the DNA was extracted once with an equal volume of 1:1 phenol:chloroform and once with an equal volume of chloroform. The DNA was then precipitated with 0.6 volumes of 5M ammonium acetate and two volumes of absolute alcohol. The pellet was washed in 70% ethanol, dried and resuspended into TE.
To isolate small fragments, the DNA was run on a 3mm thick polyacrylamide gel of appropriate percentage. The gel was stained with ethidium bromide and the bands visualised as for the agarose gel. The band was cut from the gel and put into an microfuge tube. A blue Gilson tip was used to break up the gel slice, 1ml of TE plus 0.5M ammonium acetate was added to the tube and the tube was incubated overnight at 37°C. The polyacrylamide was compacted by centrifugation and the supernatant removed. The polyacrylamide pellet was washed with 0.5 ml of TE and again centrifuged. The supernatants were pooled and the volume reduced to 400 ul by concentration with butan-2-ol. The DNA was precipitated with 2 volumes of absolute alcohol. The pellet was washed in 70% ethanol, dried and resuspended in TE.

2.5.2. "Blunt ending" restriction enzyme fragments

Restriction enzymes cut DNA to produce ends that are either blunted ended, or have 5' or 3' recessed "sticky ends". It is necessary sometimes to remove the "sticky end", for example when removing a restriction enzyme recognition site or adding linkers.

Two enzymes were used to remove "sticky ends". The 5'to 3' activity of the Klenow fragment can be used to replace the missing nucleotides from the recessed ends, ie."fill in the end". This was done as described for end-labelling except that the ³²P-alpha-dATP was omitted.

The Klenow fragment cannot fill in 5' recessed ends, but T4 DNA polymerase, which has a 3' to 5'exonuclease activity 200 times greater than that of Klenow enzyme, can be used to remove protruding 3' ends. Linearised plasmid DNA was incubated in 20 ul containing T4 DNA polymerase buffer (33mM Tris-acetate, pH 7.9, 66mM potassium acetate, 100mM MgCl₂, 0.5mM dithiothreitol, 100ug/ml BSA), containing 2mM of
each dNTP and 2.4 units of T4 DNA polymerase for 10 min at 37°C. The DNA was then extracted with an equal volume of chloroform, precipitated with 0.6 volumes ammonium acetate and 2 volumes of absolute alcohol, washed in 70% ethanol, dried and resuspended in TE.

2.5.3. Dephosphorylation of digested DNA

Calf intestinal alkaline phosphatase (CIAP) removes the 5' terminal phosphate from DNA. This is useful when ligating a DNA insert into a plasmid vector, since if the 5' terminal phosphates are missing the plasmid is unable to ligate to itself, whilst the DNA insert carrying 5' terminal phosphates is able to form a covalent bond with the dephosphorylated plasmid DNA.

Twenty minutes before the end of a restriction enzyme digest, CIAP (4 units/ug of DNA) was added to the digestion mix. At the end of the digest the CIAP was inactivated by heating to 68°C, and the DNA extracted with an equal volume of phenol:chloroform (1:1), an equal volume of chloroform, precipitated with 0.6 volumes 5M ammonium acetate and 2 volumes of absolute alcohol, washed in 70% ethanol, dried and resuspended in TE.

2.5.4. Ligation

T4 DNA ligase was used to join two ends of DNA together. This was done to either recircularise a plasmid or to insert a piece of DNA into a plasmid. To recircularise a plasmid the reaction was carried out at low concentrations of DNA to minimise intermolecular ligation, whilst to ligate a fragment into a plasmid the reaction was carried out at the highest practical concentration of DNA possible. Whenever possible the DNA to be inserted had been cut with two different
restriction enzymes, which meant that, as the plasmid would also have had to be cut with two different enzymes, it would not be able to ligate to itself. This strategy also enabled the orientation of the insert to be controlled. When this approach was not possible the plasmid vector was treated with CIAP before ligation (see Section 2.5.3).

A typical ligation reaction to insert a fragment of DNA into a plasmid would contain 25 ng of linearised vector DNA, 100 ng of insert DNA, 1/10th volume of 10x ligation buffer (0.5M Tris, pH7.5, 0.1M MgCl₂, 10mM dithiothreitol, 10mM ATP) and 0.1 units of T4 DNA ligase, in a final volume of 10 ul. The reaction was incubated at room temperature overnight. A typical recircularisation contained the same components as above, except without the insert DNA, and the reaction volume was 100 ul.

2.5.5. Ligation of linkers

Linkers were used to add new sites into the DNA. They were either bought from Boehringer/BCL or synthesised on site, (see Section 2.6). Linkers were dissolved to a concentration of 0.5mg/ml in TE. As linkers are not phosphorylated at their 5' end it was necessary to use T4 polynucleotide kinase to phosphorylate them (see section 2.4.7). The phosphorylated linkers were tested for their ability to be ligated by adding 1 ul of the kinased linkers to 10 ul of 1xligation buffer containing 5 units of T4 DNA ligase and incubating overnight at room temperature. The ligase was inactivated by heating at 70°C for 10 min. 5 ul of the ligated linkers were digested with the appropriate restriction enzyme. The unligated kinased linkers (2ul), the ligated kinased linkers (5ul) and the digested ligated kinased linkers (5ul)
were analysed by electrophoresis through a 8% polyacrylamide gel until the bromophenol blue marker had migrated 10 cm, the gel dried and autoradiographed at -70°C overnight.

The ligation mix was as described previously except that linkers were added at a higher concentration (20-50 ug/ml). The ligase was inactivated by heating at 70°C and the DNA was precipitated with 0.6 volumes of 5M ammonium acetate and 0.6 volumes of isopropanol. The DNA was resuspended and digested by the restriction enzyme for the site contained within the linker. The digest was a tenfold overdigest. The DNA was reprecipitated as before and recircularised by ligation. This removed excess linkers which may have ligated into the plasmid. The ligation mix was used to transform competent cells.

2.5.6. Colony/plaque hybridisation

On occasion it was not possible to directly select for the products of a manipulation and only a subpopulation of the colonies or plaques produced after transformation would be expected to be of interest. These were identified by colony or plaque hybridisation.

Colonies were picked from a plate and dotted onto a nitrocellulose filter placed onto the surface of a LB-agar plate, containing the appropriate antibiotics. Before use, nitrocellulose filters were pre-wet and sterilised by washing three times in dH₂O at 95°C, and dried by placing the filter between two pieces of Whatman 3MM paper. The plate was incubated overnight, after which the filter was removed and placed for 3 min onto a piece of Whatman 3MM paper soaked in 10% SDS. The filter was moved onto a piece of Whatman 3MM paper soaked in denaturing solution (0.5M NaOH, 1.5M NaCl) for 5 min, then onto a piece of Whatman 3MM paper soaked in neutralising solution (0.5M Tris-HCl, pH5.0, 3.0M NaCl) for 5 min and then finally onto a piece of
Whatman 3MM paper soaked in 2xSSC. It was then baked for 2 h in a vacuum oven at 80°C. M13 plaques could be transferred by directly placing the pre-wet filter onto the bacterial lawn, removing the filter and baking it for 2 h in a vacuum oven at 80°C.

Filters were hybridised to a probe made from the insert DNA, as described in Section 2.4.5, and autoradiographed for 5 h at -70°C. The autoradiograph was then used to locate the colonies or plaques of interest.

2.5.7. Oligonucleotide site directed mutagenesis

This was used to change the DNA sequence of a Ml3mp8hpri construct, in order to create a novel restriction enzyme site. The method was based on that of Zoller & Smith (1982). An oligonucleotide, which was complementary to the sequence to be mutated, except for the predetermined mismatches, was annealed to a single stranded DNA template and used to prime synthesis of the second strand of DNA. The addition of ligase to the reaction ensures that the synthesised strand becomes ligated to the 3' end of the oligonucleotide primer. This increases the yield of mutant phage as it prevents the entry of the endogenous DNA polymerase I enzyme and so reduces the amount of mismatch repair.

Ml3mp8hpri DNA was used to transform competent JM101 cells, which were then plated into top-agar together with 200 ul of a saturated fresh overnight culture of JM101 cells and incubated overnight at 37°C. A single plaque was used to inoculate 5ml of LB-broth plus 50 ul of a saturated fresh overnight culture of JM101 cells. The culture was incubated at 37°C whilst shaking at 200 rpm for 6 h. The culture was then aliquotted into 5 microfuge tubes and single stranded template made from each, as described in section 2.2.6. The pellets were pooled
and the OD\textsubscript{260} was measured. As 1 OD\textsubscript{260}/ml of single stranded DNA corresponds to a concentration of 20 ug/ml, the yield of the template was calculated. The DNA was precipitated and resuspended to a concentration of 1 ug/ml.

1 ul of single stranded DNA template was added to 1 ul of 10xTM buffer and 2 ul of kinased oligonucleotide (5 pmole/ul) and made up to 10 ul with dH\textsubscript{2}O in an microfuge tube. The tube was placed in a small beaker containing hot water (80°C) and allowed to cool to room temperature, which took about 40 min. To this was added 10 ul of 1xTM which contained 1 mM of each dNTP, 1 mM ATP, 10 mM dithiothreitol, 7 units of T4 DNA ligase and 2.5 units of Klenow enzyme and the mix left overnight at 14°C. This was then precipitated by the addition of 0.6 volumes of 5M ammonium acetate and 0.6 volumes of isopropanol and the precipitate resuspended in 10 ul of dH\textsubscript{2}O. This was then digested with 1 unit of S1 nuclease in 50 ul of 1xS1 nuclease buffer (28 mM NaCl, 5 mM sodium acetate, pH 4.6, 0.45 mM ZnSO\textsubscript{4}) for 5 min. The S1 nuclease was inactivated by heating at 68°C for 10 min. Triplicate 10 ul aliquots of the S1 nuclease digest were used to transform competent JM101.

The resultant plaques were picked and grown as colonies on a LB-agar plate overnight. These colonies were transferred to a filter (see section 2.5.6) and hybridised as follows. The filter was prewet in 6xSSC for 5 min and prehybridised at 67°C for 1 h in 6xSSC containing 10xDenhardt’s solution, and 0.2% SDS. The filter was rinsed in 100 ml 6xSSC, dropped onto the surface of 3 ml of 6xSSC containing the probe and left at room temperature for 1 h. The probe was made using T4 polynucleotide kinase as follows: 3 ul of oligonucleotide (5 pmole/ul) was mixed with 20 uCi of \textsuperscript{32}P-gamma-ATP, 1/10 th volume of 10xkinase buffer, 1 mM dithiothreitol, 2.5 units of T4 polynucleotide kinase made up to a final volume of 30 ul with sterile dH\textsubscript{2}O and
incubated at 37°C for 30 min. After hybridisation, the filter was washed twice with in 100ml of 6xSSC at room temperature for 1 min, wrapped in Saran wrap and autoradiographed with Fuji-RX film at -70°C for 3 h. This gave a background grid of colonies. The filter was rewashed at 10°C below the Tm expected for the oligonucleotide probe as calculated from the Wallace rule, and the filter autoradiographed as before. Colonies containing mutant phage retain the probe after the higher temperature wash whilst wildtype show a reduction in the level of signal.

Cells from the positive colonies were plated in top-agar plus 200 ul of a saturated fresh overnight culture of JM101 cells. The process was repeated again, to ensure purity of the plaques. DNA was prepared from the purified plaque phage and the presence of the mutation confirmed by sequencing.

2.6. Synthesis of oligonucleotides

Oligonucleotides were made on site and used for the EagI/KpnI linker and for oligonucleotide site directed mutagenesis. These were done by Dr. J. McCauley, using an automated DNA synthesiser (Model 381A, Applied Biosystems) working on phosphoamidite chemistry.

The oligonucleotides were synthesised with the first nucleotide bound to a column. Three washes in concentrated ammonium solution were used to remove the oligonucleotide from the column. This solution was then heated at 60°C for 5 h to remove the cyanoethyl- protecting groups. The ammonium solution was evaporated from the deprotected oligonucleotide in a vacuum dessicator (Univap). Once dry the oligonucleotide was redissolved in 0.5 ml of TE.

The oligonucleotide was purified by running on a 0.4 mm thick 20% denaturing polyacrylamide gel, containing 7M urea. The sample was
mixed with formamide dye and heated to 95°C for 5 min to remove any secondary structure before loading. When the bromophenol blue marker had migrated 2/3 of the distance, the gel was removed and sandwiched between two pieces of Saran wrap. The band corresponding to the complete oligonucleotide was visualised by UV-shadowing, i.e. by placing a fluorescense activated TLC plate under the gel and illuminating from the upper surface with UV light (300nm). As DNA absorbs UV, the DNA is seen as a dark shadow against the fluorescent background. The band corresponding to the complete oligonucleotide was cut out. The gel slice was put in a dialysis bag with about 1 ml of TE and dialysed overnight against 2 litres of TE at room temperature. The contents of the bag were put in a 10 ml syringe and pushed through a 0.45um Millex filter. The filter was washed through with 500 ul of TE and the pooled filtrate aliquoted into 5 microfuges which were dried in the vacuum dessicator. The oligonucleotide was redissolved in 500 ul of TE. The OD$_{260}$ was measured and the concentration calculated (1 OD$_{260}$ corresponds to 20ug/ml of oligonucleotide).

2.7. Computing

The graph drawing programs of Unigraph, part of the Uniras graphics package, were run on the Institute VAX 11/750 minicomputer running VMS 4.5. Unigraph is a menu driven package of programs. The line fitting programs can be accessed via the "style" menu.

Computing using sequence manipulation was carried out on the VAX 11/750 microcomputer at the AFRC computing centre (Harpenden), which was accessed via AGRENET. This computer contains both the Genbank and EMBL DNA sequence databases. These can be searched using the command "NAQ FIND".
The University of Wisconsin Genetics Computer Group (UWCG) sequence analysis software package (Version 5.0) was used to manipulate DNA sequences (Devereux et al, 1984). The following programs were used:

i) **FETCH**  this copies from a database to a directory.

ii) **MAP**  this displays both strands of a DNA sequence, with a restriction map shown above and the possible protein translations shown below.

iii) **SEQED**  this is an interactive editor for entering or editing DNA sequences. It was used to assemble parts of the neo and hprt DNA sequences into new genetic constructs. Sequence data can be entered by the file or directly from the keyboard.

iv) **FIND**  this looks for patterns in a DNA sequence. It can be asked to allow for mismatches.

All computing was carried out using a Pericom terminal.
CHAPTER 3

CHARACTERISING THE PLASMIDS USED IN THE STUDY
3.1. Introduction

The basis of this project is the design of a plasmid which will allow the direct selection of a cell in which integration has been targeted into the hprt gene. This plasmid has three components:

a) a neo gene which is in a form that is expressible in mammalian cells;

b) a stretch of DNA that is homologous to the hprt gene;

c) a bacterial plasmid component which allows manipulation of the construct in E. coli, but which is suitable for transfection into mammalian cells.

In this chapter I will describe the preliminary studies to establish that these components were suitable for the construction of the plasmid which is described in Chapter 6.

The plasmid pdBPVMntneo replicates in both E. coli and mammalian cells (Law et al, 1983). In addition to the 69% subgenomic fragment of BPV, it possesses a neo gene in an expression cassette which uses the mouse metallothionen I promoter (Mnt) and the SV40 early region small t antigen splicing signals and 3' transcriptional processing signals. The bacterial plasmid component is pML2d, a plasmid derived from pBR322 by deletion of the DNA sequences that interfere with replication in mammalian cells. In this study, a derivative of this plasmid, from which the BPV component had been removed, was used. This plasmid was termed pML2dMntneo, and was a gift from R. Meehan (MRC-CAPCU, Edinburgh).

As the neo gene is a bacterial gene that confers resistance to kanamycin, I have investigated whether the neo gene is expressed in
E. coli, and if so, which region contains the bacterial promoter for this gene and what are some of its characteristics (Sections 3.2 and 3.3).

Since pML2dMntneo no longer contained the BPV component it would not be expected to replicate extrachromosomally in mammalian cells, and so transformed colonies, that were resistant to G418, would be found to have the plasmid integrated into the genome. The characteristics of transformed cells resulting from transfections using this plasmid are discussed in Section 3.4.

In addition to the neo and plasmid components, a source of DNA homologous to the hprt gene was required. I used the human hprt cDNA. Two plasmids, each containing independently derived cDNAs of the human hprt gene, **p4A8** (Jolly et al., 1983) and **pHPT30** (Brennand et al., 1983) were obtained. The plasmid p4A8 was a gift from T. Friedmann (University of California, San Diego) and pHPT30 from D. Melton (University of Edinburgh). These are referred to as Friedmann’s and Caskey’s isolates respectively. Friedmann’s isolate was cloned into the vector of Okayama and Berg (1983) which enables the cDNA to be directly expressed in mammalian cells (see Section 1.4.3). For this reason, this isolate was used in transfection and microinjection experiments (Section 3.6). The hprt sequence of this isolate had been published and so it was used to construct the plasmid for targetted integration. In order to be confident that the cDNA obtained did not deviate from the published sequence, it was checked by sequencing. As a comparison the majority of Caskey’s isolate was also sequenced (Section 3.5).
Figure 3.1 Mapping of pML2dMmtneo

A: The map of pML2dMmtneo, showing HindIII, EcoRI, KpnI, BglII, EagI, PstI and BamHI sites. The bacterial plasmid vector sequences are shown as a single line, inserts (Mmt promoter, neo and SV40 small t antigen splice and 3' transcriptional processing signals) shown as blocks.

B: Double digests of pML2dMmtneo were separated on a 0.7% agarose gel, stained with ethidium bromide and photographed. The digests were as follows:

a) BglII and KpnI
b) BamHI and PstI
c) EcoRI and KpnI
d) BglII and BamHI
e) BglII and EcoRI
f) EcoRI and PstI
g) EcoRI (alone)
h) BamHI and EcoRI
i) BamHI and KpnI
j) PstI (alone)
(EagI digests are not shown)
M, HindIII digests of Lambda DNA, (Sizes shown are in Kb).

C: End-labelled fragments of EcoRI, EcoRI/HindIII and HindIII digests of pML2dMmtneo separated on a 6.0% polyacrylamide gel and autoradiographed.

This shows fragments of approximately 60bp and 30bp, and is consistent with the presence of an inverted repeat of the small EcoRI/HindIII fragment of pBR322 DNA.
M, end-labelled HinfI digested M13mp8 DNA.
M', end-labelled HinfI digested pGEM2 DNA.
(Sizes shown are in bp).
3.2. The plasmid pML2dMmtneo

The plasmid pML2dMmtneo was used to transform the bacterial host HB101 and the transformants were selected on kanamycin plates. This plasmid was found to confer resistance to kanamycin at a concentration of 50 μg/ml. The restriction enzyme map of the plasmid produced from these colonies is consistent with that expected, (Figure 3.1a,b). The sites for KpnI and EagI have been mapped in addition to those shown on the map by Law et al (1983). An inverted duplication of the 29 bp. EcoRI to HindIII fragment of pBR322 at the junction of the Mmt and pML2d components of pdBpVlmlmtneo was described by Law et al (1983). pML2dMmtneo was digested with EcoRI or both EcoRI and HindIII, the resulting fragments were end-labelled with $^{32}$P-dATP and separated on a 6% polyacrylamide gel. The appropriate bands could be seen demonstrating that the duplication is retained in pML2dMmtneo, (Figure 3.1c).

As the plasmid is able to confer kanamycin resistance to E. coli, where is the bacterial promoter? The promoter for the tetracycline resistance gene of pBR322 is within the 29 bp EcoRI-HindIII fragment. It was important to establish whether this region contained the functional bacterial promoter for the expression of the neo gene, because, as it is flanked by EcoRI sites, it would be removed by EcoRI digestion and therefore have had consequences for the way in which the plasmid could have been manipulated. The duplicated region was removed by digestion with EcoRI and recircularisation by ligation at a low DNA concentration. The products of this ligation were then used separately to transform E. coli which were then plated onto kanamycin and ampicillin plates. Approximately equal transformation frequencies were obtained when selected on both kanamycin and ampicillin. To ensure that the duplicated sequence had in fact been removed, plasmid minipreps were made from a number of kanamycin-resistant colonies and
Figure 3.2  Removal of the EcoRI/HindIII inverted repeat from pML2dMmtneo

For details of the experiment see Section 3.2.

The EcoRI (E) and HindIII (H) of minipreps from five plasmids, made as described in text, were separated on 0.7% agarose gel, stained with ethidium bromide and photographed. Each EcoRI digest linearised the plasmid to produce a 6.7 Kb band, whilst HindIII did not digest the plasmid. Each HindIII digest produced a band corresponding in size to that of HindIII digested pUC19 (marked with arrow). This demonstrated that the pUC19 DNA mixed in each HindIII digest was cut by the enzyme.

The tracks pUC19 and Un contain HindIII digested pUC19 and undigested pML2dMmtneo respectively.

The track M contains a HindIII digest of Lambda DNA.
digested with EcoRI or HindIII. In all cases the plasmids could be
digested by EcoRI but not HindIII, (Figure 3.2). Included in each of
the HindIII digests was the plasmid pUC19, which contains a single
HindIII site. In all cases the pUC19 was digested by HindIII,
demonstrating that HindIII did not digest the miniprep DNA because no
HindIII site was present, and not because a contaminant in the
miniprep DNA inhibited the enzyme. As the HindIII site is situated
within the duplicated region, it must have been removed from each of
the plasmids. Loss of the duplicated region therefore did not remove
the bacterial promoter for the neo gene.

The Mmt promoter was removed from pML2dMmtneo by digestion with
EcoRI and BglII, and replaced by a BamHI linker. The resulting
plasmid, designated p(Bam)neo, did not confer kanamycin resistance to
E. coli at a concentration of 50 ug/ml. This suggested that the
functional bacterial promoter of the neo gene lay within the Mmt
promoter region.

A simple assay, based on the one described by Reiss et al (1984),
was used to compare the expression of the neo gene from the putative
promoter in the Mmt region with that from the known bacterial
promoter. The maximum concentration of kanamycin to which a plasmid
confers resistance is proportional to how efficiently the neo gene is
expressed. Bacterial colonies, transformed with either the plasmid
pML2dMmtneo or p(Bam)neo, were replica plated onto plates containing
kanamycin at concentrations varying from 0 to 500 ug/ml. The results
are shown in Figure 3.3a. pML2dMmtneo had an maximum level of
resistance of 200 ug/ml, this compares poorly with the plasmid pKm2
which expresses neo from the bacterial LacUV5 promoter and is
resistant to greater than 1000 ug/ml kanamycin (Reiss et al, 1984).
Removal of the Mmt promoter did not completely abolish the expression
Figure 3.3  Expression of the neo gene in E. coli

A: Histogram of the maximum level of kanamycin resistance of E. coli (strains DH5 and HB101) and E. coli HB101 after transformation with p(Bam)neo and pML2dMmtneo. This was measured using the method of Reiss et al (1984), and is described in the text (see Section 3.2).

B: This shows the sequences of the putative bacterial gene promoters identified using the Find program of the Wisconsin (UWGCG) sequence analysis package, and their respective homology scores (see Section 3.3). The nucleotides in common with the consensus sequence are boxed. As a comparison, the promoter of the tetracycline resistance gene from pBR322 is also shown.
HOMOLOGY SCORE

Consensus sequence

TGGAGAT ----------------------- TATAAT

tet promoter

ATTCTGAGTTATGATAGATACGATCC

66%

Mtn promoter

GCAAGGGCTGCTGCGGATCTGCAACAGATCCAGAG

41%

Cryptic promoter

TGGCGGAGGCCGGAGCTAAGAAAGCAT

0.32% 49%
of neo, as p(Bam)neo conferred resistance to a maximum of 15 ug/ml. This level of expression was assumed to result from a cryptic promoter in the plasmid sequences 5' of the Mmt promoter that became aligned with the neo gene after removal of the Mmt promoter. Bacteria with no plasmid showed no basal level of resistance in this assay.

3.3. A computer search for neo promoters

The possible existence of a bacterial promoter within the Mmt region and a cryptic promoter 5' of the Mmt region prompted a search for possible candidates within these DNA sequences. The DNA sequences of pBR322 and the 280 bp 5' portion of the Mmt promoter were obtained from the Genbank data base using the FETCH program of the Wisconsin (UWGCG) sequence analysis package. Bacterial promoters were searched for in the following way. The consensus bacterial promoter has the following sequence:

```
---TTGACAGT-------------TATAAT---
  -35             -10
```

ie. two consensus elements spaced apart by a gap of 15 to 21 base pairs (reviewed by McClure, 1985). Each of these components were taken separately and the program FIND used to search the pBR322 and Mmt sequences. As the promoter sequence is a consensus, and no known bacterial promoter is identical to it, three mismatches were allowed in the search. This identified a long list of possible sites in the pBR322 sequence and a number for the Mmt sequence. In order to reduce
the number of possible sites only those possessing the following criteria were accepted:

i) -35 elements that contained one or more nucleotides of the sequence TTG____;

ii) -10 elements that contained two or more nucleotides of the sequence TAT_T.

These reduced lists were superimposed to make a final group which consisted of a set of sequences with the -35 element between 10 and 30 bp upstream of the -10 element. This identified a single putative bacterial promoter sequence within the Mnt promoter and two sequences in a region of the pBR322 sequence, with a common -10 element, that could be the cryptic promoters of the plasmid p(Bam)neo (see Figure 3.3b).

In order to evaluate the significance of these putative promoter sequences, an adaptation of the formula from the computer program, described by Mulligan et al (1984) for searching for bacterial promoter sequences, was used to give a score of the homology of these sequences to bacterial promoter sequences. In this program a weight table, based on the occurrence of each nucleotide of 112 promoter sequences, is used to score each nucleotide in a region that extends across the putative promoter. The spacing is also given a score and a baseline corresponding to a quarter of the maximum score is subtracted in order to eliminate the effect of chance similarities. The following formula is used to express the score for the promoter under test as a percentage of the maximum possible score:

\[
\text{HOMOLOGY SCORE} = \frac{100(\text{nucleotide score + spacing score - baseline score})}{(\text{maximum score - baseline score})}
\]
Figure 3.3b, shows the homology scores of these promoters and that of the tetracycline (tet) promoter. The putative Mmt bacterial promoter scores 25% less than that of the tet promoter, which is consistent with the reduced expression from the Mmt promoter in comparison with an authentic bacterial promoter.

Of the two putative cryptic promoters, one, marked (ii) in Figure 3.3b, has a spacing of less than 15 nucleotides and therefore may not be functional. The other putative promoter has a score of 32% and a spacing of one nucleotide greater than the consensus of 17 nucleotides. Although the reduced level of expression is consistent with the reduced homology score of this promoter, a 9% difference in the homology score corresponds to a reduction in the level of expression of the neo gene to 7.5% of that observed from Mmt promoter. This could be simply explained if the relationship between homology score and level of expression is not linear, so that a promoter with a low homology score functions disproportionately poorly. It is possible however that p(Bam)neo may possess sequences between the promoter and the neo gene which reduce the level of expression.

Bacterial mRNAs possess signals that are required for binding of the ribosome thereby allowing efficient translation. Stormo et al (1982) produced a weight matrix (W101) which covers 60 nucleotides upstream and 40 nucleotides downstream of the initiation codon. This matrix proves to be the best means of determining functional initiation codons. It covers a length of sequence that extends beyond the region known to be in direct contact with the ribosome during initiation of translation. Why this extra sequence is required is not known. It may be an artefact of the way the matrix was produced (Stormo et al, or mean that this extra sequence is also necessary for
Figure 3.4  The change in the transfection frequency of A9 and A2-4 cells with increasing amounts of calcium phosphate precipitate containing pML2dMMntneo

Plot of the number colonies obtained after transfection of A9 or A2-4 cells with different amounts of precipitate, ranging from the equivalent of 0.025 to 0.40 ug DNA per surface area (cm$^2$) on which the cells were growing. See Section 3.4 for experimental details.

Bars show the Standard Error of the Mean.
Amount of precipitate (µgDNA/cm²)

No. of colonies/10⁵ cells transfected

- • A2-4
- ★ A9
translation, eg. through secondary structure effects. In the plasmids tested here only 36 of the upstream nucleotides are derived from the neo gene, the rest are from either the Mmt or pBR322 components. These sequences therefore may have an influence on the efficiency of translation. This matrix gives the following scores: neo plus Mmt sequence, 17; neo plus cryptic promoter, 126 and neo in the context of Tn5 sequence (ie. wildtype), 70, a score greater than 2 identifies all known gene starts. Within the limitations of this test, it would appear that the influence of the sequences between the promoter and the neo gene is not significant, and therefore, at lower values the homology score is not linearly related to the level of expression.

3.4. Transfection of the plasmid pML2dMmtneo

To ensure that if homologous recombination occurred it could only take place between the plasmid and chromosomal DNA, it was necessary to transf ect the plasmid DNA without using carrier DNA. Confirmation was needed that pML2dMmtneo could be used in a carrier-free system to transform the A9 and A2–4 cells to G418 resistance. For a high transfection frequency in the absence of carrier DNA, the precipitate must be formed with the plasmid DNA at a concentration of 20 ug/ml (Huttner et al, 1981). To avoid unnecessary use of DNA it was important to know the relationship between the amount of precipitate and the transfection frequency.

Dishes containing $10^5$ A9 or A2–4 cells were transfected with varying amounts of precipitate, and the number of colonies counted after 2 weeks on G418 selection. The results are shown in Figure 3.4. The transfection frequency increased linearly until the amount of precipitate reached a value of 0.15 ug DNA/cm$^2$, further increase in
amount of precipitate did not increase the transfection frequency and possibly caused the frequency to drop. From the shape of the curve it appeared that an optimum was reached. The large variation between the transfection frequencies of individual experiments, when carrying out transfections with larger amounts of precipitate, made it difficult to compare different points on the curve. Analysis of the data using the Student t-test to compare each frequency, at the other amounts of precipitate, with those at 0.15 ug DNA/cm$^2$, showed that only the transfection frequencies obtained with 0.02 and 0.05 ug DNA/cm$^2$ were significantly different. It was therefore only possible to conclude that the transfection frequency was limited by the amount of precipitate when less than the equivalent of 0.1 ug DNA/cm$^2$ was put onto the cells.

The variation in transfection frequency is a problem inherent in the calcium phosphate precipitation method. As seen in Figure 3.4, there is a large variation in the transfection frequency even using a single precipitate. This makes it difficult to measure accurately the transfection frequency of a particular plasmid, but taking the mean transfection frequency for the three values beyond the saturation point in Figure 3.4, frequencies of $1.02 \times 10^3$ for the A9 cells and $1.11 \times 10^4$ for the A2-4 cell line were obtained.

It is interesting to ask whether the number of copies of plasmid integrated increases with increasing amounts of precipitate. This was investigated by picking colonies that resulted from transfections with 0.05, 0.1, 0.15 and 0.4 ug DNA/cm$^2$ of precipitate. DNA from each colony was digested with XbaI, an enzyme that does not cut within the plasmid. Each site of integration of plasmid DNA would be expected to produce a single band comprising of a concatamer of plasmid molecules linked to additional chromosomal DNA of random sizes, which derives
10 ug of each DNA was digested with BamHI (panel A) or XbaI (panel B), separated by electrophoresis, blotted onto Biotrace-RP membrane, hybridised to $^{32}$P labelled pML2dMmtneo DNA and autoradiographed for 3 days.

**A:** Conventional electrophoresis was used to separate DNA on a 0.7% agarose gel. DNA from colonies picked after transfection with 0.05, 0.1, 0.15 and 0.4 ug DNA/cm$^2$ of precipitate was analysed, see Section 3.4 for experimental details.

The 6.7 Kb band indicative of a head-to-tail concatamer of plasmid is indicated.

The tracks labelled 0.05-3' and 0.15-2' are the a 1 day exposure of tracks 0.05-3 and 0.15-2.

A number of bands are present in the A9 DNA. These appear to be due to contamination with marker DNA during loading of the gel and a plasmid contaminant present in all DNAs (*).

The gel included copy controls calculated to be the equivalent of 1, 10 and 100 copies of the BamHI digested pML2dMmtneo DNA per cell. The extra bands are due to partial digestion.

The track M contains a HindIII digest of Lambda DNA.

**B:** DNAs from the same colonies as A, plus some additional lines, were separated by FIGE.

The track marked 0.4-2' is a 1 day exposure of the autoradiograph and shows the band in track 0.4-2 more clearly.

Track M contains a concatameric series generated by ligation of EcoRI digested pBR328/tk/Sat.

All sizes are in Kb.
Table 3.1 Estimation of the copy number of the plasmid pML2dMmtneo following transfection with differing amounts of calcium phosphate precipitate

<table>
<thead>
<tr>
<th>Amount of precipitate&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Size (Kb)&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Maximum no. of copies&lt;sup&gt;d&lt;/sup&gt;</th>
<th>Estimate from BamHI digest&lt;sup&gt;e&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05 (1)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>60, 35</td>
<td>8, 5</td>
<td>13</td>
</tr>
<tr>
<td>0.05 (2)</td>
<td>38</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>0.05 (3)</td>
<td>S</td>
<td>S</td>
<td>70</td>
</tr>
<tr>
<td>0.1 (1)</td>
<td>25</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>0.1 (2)</td>
<td>33</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>0.15 (1)</td>
<td>16</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>0.15 (2)</td>
<td>S</td>
<td>S</td>
<td>100</td>
</tr>
<tr>
<td>0.4 (1)</td>
<td>23</td>
<td>3</td>
<td>ND.</td>
</tr>
<tr>
<td>0.4 (2)</td>
<td>20</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

<sup>a</sup> The unit for the amount of precipitate is ug DNA/cm².

<sup>b</sup> These numbers refer to the numbering of the DNAs on Figure 3.5a, eg. 0.5 (1) corresponds to track 0.5-1.

<sup>c</sup> Sizes were estimated relative to the plasmid markers (see Materials and Methods).

<sup>d</sup> The maximum copy number was given by the first integer below the figure obtained by dividing the size of the band by 6.7 (ie. the size of a single plasmid).

<sup>e</sup> Estimated copy number from the intensity of the bands relative to the copy number controls.

S refers to those lines where a smear rather than a discrete band was obtained (see text).

ND., not calculated.
from the sequences flanking the site of integration. It was expected that several copies of plasmid would integrate at each site, so the digests were separated by FIGE under conditions covering the maximum range of sizes. The DNA from the gel was transferred to a membrane and hybridised to a probe made from the plasmid. Figure 3.5b shows the autoradiograph of this gel, and the sizes of the bands and the maximum number of copies they could contain are shown in Table 3.1.

Figure 3.5b shows that this approach can be used to estimate the number of copies integrated. Most of the DNAs gave discrete bands, the sizes of which indicate the maximum number of copies of plasmid that could have integrated. This ranges from 2 to 8 copies. As XbaI is an enzyme that cuts at a hexanucleotide site it is expected to cut genomic DNA to a mean length of 4 Kb. In most cases the bands produced on the autoradiograph contained DNA of approximately this length after the size of the maximum amount of plasmid DNA has been accounted for. The difference between the intensities of the bands may reflect a smaller copy number than suggested from the maximum possible. Two tracks show smears that hybridise the probe. The smearing in both cases covers a size range from approximately 20 Kb to greater than 100 Kb. The reason for these smears is not clear.

The above observations were confirmed by digesting the same DNAs with BamHI, which cuts at a unique site within the plasmid, and analysing the digests on a conventional 0.7% agarose gel (Figure 3.5a). If every plasmid had become joined in a head-to-tail concatamer then a single 6.7 Kb band would be seen with the intensity of the signal being proportional to the copy number. In addition, there would be two unique junction fragments for each site of integration. The 6.7 Kb band is present in all but two tracks. The intensities of the 6.7
Kb bands for each DNA are consistent with the conclusions drawn from the FIGE gel, with a few exceptions.

1) The DNA from line 0.05-2, gave an apparent copy number of up to five plasmids with FIGE, however only gave a single high molecular weight band after the BamHI digestion and conventional electrophoresis. It could be that only a fragment of the original plasmid integrated; the plasmid lost its BamHI site before concatamerisation; or that the plasmid integrated into a region with lower than average abundance of BamHI and XbaI sites. The reduced intensity of signal from the single band with the XbaI digests is consistent with only a small amount of plasmid DNA integrated.

2) The BamHI digest of 0.15-1 did not produce a band of equivalent size to the linear plasmid, which can be explained by suggesting that this line has only integrated a single copy of the plasmid.

The DNAs, 0.05-3 and 0.15-2, that gave smears on the FIGE were digested by BamHI to give discrete bands. From the band intensity it is clear that both DNAs contain a large number of copies of the plasmid, estimated to be approximately 70 and 100 copies, respectively. An explanation for the smearing that was observed following FIGE could be that these two transformants could have produced one or a few very large bands, but that the DNA was degraded to allow the resolution of intact bands. The possibility that the plasmids in these cells formed an extrachromosomal structure that runs aberrantly on the FIGE cannot be eliminated. Taking all of these observations into consideration, there is no correlation between an increase in the amount of precipitate and the number of copies integrated.

The BamHI digest also gave information on the structure of the plasmid molecules within the integration site. It is clear that, although the 6.7 Kb band is present, in most cases there are
Figure 3.6  Map of p4aA8

A: Map of p4aA8 showing ClaI, HindIII, XhoI, BamHI, and PstI sites. Plasmid vector sequences are shown as a single line and the inserts (SV40 early gene promoter, human hyp cDNA and SV40 late gene polyadenylation signals) are shown as blocks.

B: The following digests of p4aA8 were separated on a 1% agarose gel, stained with ethidium bromide and photographed:
   a) ClaI (alone)
   b) BamHI and ClaI
   c) BamHI (alone)
   d) HindIII (alone)
   e) HindIII and ClaI
   f) HindIII and PstI
   g) PstI (alone)

The tracks marked M contain a HindIII digest of Lambda DNA, (sizes shown in Kb).
A

Clai/HindIII (4.4/0)
PatI (3.8)
BamHI (0.3)
PatI (0.4)
Early SV40 promoter
XhoI (0.6)
XhoI/BamHI (1.6)
HindIII (1.0)
poly A

B

M a b c d e f g M

23.1-
9.42-
6.56-
4.38-
2.32-
2.02-
additional bands, some of which cannot be junction fragments. These were due to rearrangement of some of the copies of integrated plasmids within the integration site, and presumably resulted from nonhomologous recombination events occurring either before or after the integration event. It was noted that in the lines 0.5-3 and 1.5-2, where there were large numbers of plasmid present, that there were more rearranged forms. This suggests that the amount of rearrangement is proportional to the amount of plasmid present. With shorter autoradiographic exposures of the line 0.5-3, there are some rearranged bands present with higher intensities than others. This suggests that more than one copy of some rearranged plasmids are present, this may be due to a process of amplification, either before or after integration.

3.5 Hprt sequence

The restriction enzyme map of the plasmid p4aA8 was found to be consistent with available published information (Figure 3.6). In order to sequence the insert it was necessary to subclone it into M13 vectors. The subclones and the number of nucleotides of the sequence obtained are illustrated in Figure 3.7. The hprt insert was removed from p4aA8 as a PstI-BamHI fragment and ligated into PstI-BamHI digested M13mp8 or mp19. Subclones 8F and 19F gave 300 nucleotides of the 5’ hprt sequence and 231 nucleotides of the 3’ hprt sequence respectively. To obtain the sequence from internal regions of cDNA, the unique HindIII and HincII sites within the hprt cDNA were used to subclone the insert further as follows:

1) by digesting the replicating form (RF) of 8F with HindIII and then recircularising it, removing the 573 bp 5’ hprt HindIII fragment to create the subclone 8FHin;
This illustrates the location, direction and lengths of the sequences obtained with respect to the published sequence (Jolly et al, 1983). Each name refers to the M13 phage subclones that contained the fragments of either Friedmann’s (F) or Caskey’s (C) human cDNA isolate. The constructions are described in Section 3.5.
ii) by digesting RF of 19F with HindIII and HincII, filling-in the ends and then recircularising it, removing the 3' 916 bp HincII fragment to create the subclone 19Hc;

iii) constructing the subclone 8FHiPs by subcloning the 5' PstI-HindIII hprt fragment of p4aA8 into M13mp8 cut with HindIII and PstI.

The resulting sequence covered one or other of the strands of the hprt insert except for 2 gaps; a 7 bp gap from nucleotide 560 to 573 and a 200 bp gap from nucleotide 870 to 1070, (the numbering of the nucleotide corresponds to that used on the published sequence). The region of the hprt cDNA eventually used to construct the plasmid for targeted integration was completely sequenced and no differences were found from the same region in the published sequence.

Caskey’s isolate is a not a full length cDNA and consists of an insert equivalent to the 5' 952 bp of Friedmann’s isolate. It was cloned as PstI fragment and so insertion of the fragment into PstI cut M13mp8 produced subclones of either orientation. In Figure 3.7, are shown the subclones and number of nucleotides sequenced of Caskey’s isolate. Subclone C3 gave 250 bp of sequence from position 50 to 300 whilst subclone C4 gave 240 bp from the 3' end from position 800 to 910. As with Friedmann’s isolate it was necessary to further subclone the cDNA to complete the sequence. To achieve this for both orientations it was possible to use a HindIII digest followed by recircularisation to produce subclones from which was obtained sequence from position 300 to 490 and 600 to 840. This gave 750 nucleotides of the total 952, ie. 79%. All the sequence obtained from Caskey’s isolate was identical to that of Friedmann’s isolate.
Figure 3.8 Analysis of DNA from HAT-resistant colonies produced by microinjection of A9 cells with p4aA8

10 ug of each DNA was digested with EcoRI (panel A) or BamHI (panel B), separated by electrophoresis, blotted onto Biotrace-RP membrane, hybridised to $^{32}$P labelled p4aA8 and autoradiographed for 3 days.

A: EcoRI digests of DNA from three microinjectant lines were separated on a 0.5% agarose gel.

B: BamHI digests of DNA from the same three microinjectant lines were separated on a 1.0% agarose gel.

The marker sizes are in Kb.
3.6. Introduction of p4aA8 into A9 cells

To establish that the plasmid p4aA8 could express a functional hpert protein, it was introduced into A9 cells by microinjection. Initially cells were injected with DNA at a concentration of 10 µg/ml (this is equivalent to injecting between 10-50 copies into each nucleus). HAT-resistant colonies were produced at a mean frequency of 3.5 colonies/100 cells injected. In order to attempt to produce a cell that had a single integrated copy of the plasmid p4aA8, A9 cells were injected with DNA at a concentration of 1 µg/ml. This produced HAT-resistant colonies at a mean frequency of 0.5 colonies/100 cells injected. Some of these colonies were picked and grown to produce lines which were named M13, M5:2 and M32:2. M13 was produced by microinjection of one tenth the usual DNA concentration.

The DNA from these lines were analysed, by blotting and hybridisation with p4aA8 plasmid DNA, to determine the number of copies of the plasmid integrated. In Figure 3.8a, the DNA was digested with EcoRI which does not cut within the plasmid. As with the XbaI digest of the cells transfected with pML2Mmtneo (Figure 3.5b), the size of the band which hybridises the probe is related to the number of copies of the plasmid integrated. The DNA from M13 contained a single band 9.6 Kb in size. As the plasmid p4aA8 is 4.4 Kb in size, this band could contain one or two copies. The DNAs from M5:2 and M32:2 contain up to four and greater than ten copies respectively. In M5:2 the plasmid was integrated into a single site and in M32:2 was integrated into three sites. Figure 3.8b shows BamHI digests of these DNAs. BamHI cuts the plasmid at two sites to give a 1.4 Kb and a 3.0 Kb fragment. The 1.4 Kb fragment contains the hpert gene, so the plasmid would not be expected to integrate into the genome through a site within this fragment, as this would disrupt the expression of the
hprt gene. If a single plasmid molecule had integrated then this 1.4 Kb and two novel junction fragments would be formed. If more than one plasmid had integrated as a head-to-tail concatamer then a 3.0 Kb fragment would be seen. The M13 DNA gave three faintly hybridising fragments, one of which is 1.4 Kb and neither of the others are 3.0 Kb. This suggests that in the line M13 a single plasmid has integrated, and so suggesting that this line contains a single hprt gene. The other DNAs showed the presence of a 3.0 Kb fragment in addition to a 1.4 Kb fragment and two junction fragments. This is consistent with integration of more than one copy of the plasmid arranged in a head-to-tail concatamer at a single site.

The number of copies of plasmid integrated after microinjection was similar to that integrated in the majority of cases after calcium phosphate precipitation. The results of microinjection differs in that there is little rearrangement in comparison with the rearrangements that were seen when pML2dMmtneo was transfected into A9 cells. This difference between microinjection and transfection by the calcium phosphate precipitation method has been noted previously (Simons, 1985).

3.7. Discussion

In this chapter are presented the investigations made on the plasmids pML2dMmtneo and p4aA8, which established their suitability for future experiments.

Plasmid pML2dMmtneo confers kanamycin resistance to bacteria and the bacterial element responsible for the expression of the *neo* gene has been localised to the *Mmt* region. Also identified is a possible cryptic promoter in the plasmid sequence 5' of the *Mmt* region, which gives low levels of expression of the *neo* gene.
The observation that a mammalian promoter can function in *E. coli* is not unique. A region within the LTR of avian tumour virus (ATV) DNA has also been shown to express the *neo* gene in *E. coli* (Mitsialis et al, 1981). The Mmt promoter has a TATA element within its promoter which acts as the -10 region of the bacterial promoter. Many eukaryotic promoters contain such a sequence. If they possess an appropriately placed upstream sequence that is homologous to the -35 region of the bacterial promoter, these also may function in bacteria. The fact that the Mmt promoter is a promoter in both systems proved useful for the manipulations described in Chapter 6.

In Section 3.4, it was shown that the plasmid pML2dMmtneo efficiently transforms both A9 and A2-4 cell lines. The Mmt promoter expresses *neo* at a level sufficient for selection, and it is possible to isolate colonies where only a few copies of the plasmid have integrated. No induction of the Mmt promoter is required for expression of the *neo* gene. It was shown that the transfection frequency obtained with this plasmid was only limited when the amount of precipitate was low, and that the amount of precipitate put onto the cells was not related to the number of copies of plasmid that were integrated into stable transformants. This gave a figure for an optimum amount of precipitate to get the maximum transfection frequency with economy of the amount plasmid DNA used. By expressing the amount as ug DNA/cm², it is easy to adjust the amounts for different dimensions of surface on which the cells were grown. Further transfections generally used 0.2 ug DNA/cm², equivalent to 15.0 ug of DNA in a 75 cm² flask, to ensure that the cells were saturated with the precipitate.

This saturation effect is also seen with transfections using carrier DNA (Graham et al, 1980; Cartier et al, 1987). These cases
differed from the transfections described here as the amount of precipitate remained constant whilst the proportion of plasmid was varied. This suggests that it is the way the cell processes the plasmid DNA, rather than the amount put onto the cells, that determines the transfection frequency. This is consistent with the observations discussed in the Introduction (Section 1.2).

Within a single integrated structure plasmid molecules were in one of two forms. The majority were arranged as head-to-tail concatamers. This is thought to be produced by homologous recombination between plasmids shortly after entry into the nucleus (Folger et al, 1982). Some of the plasmid molecules became rearranged, which may have been due to either a combination of degradation and ligation, or some more complex nonhomologous recombination process. These rearrangements are similar to those observed with other plasmids introduced into cells by the calcium phosphate precipitation method (Huttner et al, 1981; Simons, 1985). In addition to rearrangements, there is some evidence that a form of amplification may have occurred. It may be that the plasmid can undergo limited replication before integration. Alternatively, regions of the integrated plasmid concatamer may become amplified. Amplification of plasmid DNA has been investigated by Roberts et al (1983). It was suggested that integrated plasmid DNA could become amplified in a single step to produce a 20-40 fold increase in the copy number. The amplification process appeared to affect some regions more than others. It may be that after transfection, a concatamer of plasmids happen to integrate into a region of the genome which is prone to amplification, eg. into a region which undergoes an abnormal amount of replication. Cells that contain large numbers of integrated plasmids that show signs of amplification, may be a means of identifying and isolating regions of
high replication activity, and therefore perhaps even mammalian "replication origins".

The two isolates of the hpmt cDNA are shown to be identical in sequence, over the regions I sequenced. The cDNA sequences of the hpmt genes from human, mouse and Chinese hamster are known to be conserved, at least 95% homologous at the amino acid level (Konecki et al., 1982). It is not surprising that both human isolates are identical. It therefore seems reasonable to assume that there will be no difference between Friedmann's isolate and the exon sequences of the hpmt gene of the A2-4 cells into which integration through homologous recombination will be sought.

It was verified that the plasmid p4aA8 could transform A9 cells to HAT resistance. By using microinjection rather than transfection by the calcium phosphate precipitation method, it was possible to reduce the number of copies injected to a quantity of plasmid small enough to generate a transformed cell line that contains only a single integrated plasmid molecule. This cell line, M13, was used in later experiments. The structure of the plasmid DNA after microinjection was more simple than after transfection, and consistent with the idea that all of the plasmid DNA formed a structure arranged as a head-to-tail concatamer.

To conclude, the plasmids pML2dMmtneo and p4aA8 were suitable to introduce into mammalian cells and for the construction of the fusion gene that was to be used for targetted integration into the chromosomal hpmt gene.
4.1. Introduction

The scheme outlined in the Introduction (Section 1.11) is designed to select for those cells that have integrated a plasmid into the \textit{hprt} gene. These cells would be expected to be both \textit{6TG} and \textit{G418}-resistant, as they will have simultaneously lost the \textit{hprt} gene and begun to express the \textit{neo} gene under the control of the \textit{hprt} promoter. It is possible that this scheme will also indentify those cells which have spontaneously lost the \textit{hprt} gene and simultaneously integrated the promoterless \textit{neo} gene into another location at which it is expressed. In the work discussed in this chapter, the individual frequencies of these events are estimated, in order to assess whether their occurrence will interfere with the isolation of those cells where plasmid integrated has been targetted into the \textit{hprt} gene.

4.2. The stability of \textit{hprt} gene expression in the cell lines A2-4, M13 and 4a5.

Both A2-4 and M13 cell lines are used in the targetted integration experiments. They would be expected to lose their \textit{hprt} genes, and so revert to \textit{6TG} resistance, by different mechanisms. The line A2-4, is a mouse/human hybrid cell that contains a single human X-chromosome in an A9 parent cell. Human chromosomes are lost from hybrid cells at a high frequency (Weiss & Green, 1967), so the rate of loss of the X-chromosome was expected to be the major influence on the rate at which the population of A2-4 cells reverted to \textit{6TG} resistance. The M13 cell line was made by microinjecting the plasmid p4aA8 into an A9 cell and shown to contain a single integrated copy of the plasmid (see Section 3.6). M13 was expected to revert to \textit{6TG} resistance through loss or inactivation of the integrated \textit{hprt} gene, which was expected to occur less frequently than the loss of the X-chromosome.
from the A2-4 hybrid cells. The cell line 4A5 was isolated after transfection of A9 cells with the plasmid p4aA8, and was estimated to contain 6 or 7 copies of the plasmid which had integrated at two different sites. This would also be expected to revert to 6TG resistance due to loss or inactivation of the gene, however these would be expected to arise less frequently than in M13, as at least two events would be expected to be necessary to remove all of the functional copies of the gene.

A stock of each cell line was maintained on non-selective medium. Previous to this, each line had been maintained on HAT medium. Each culture was sampled at various times by taking a fixed number of cells and plating them in medium containing 6TG. The number of resultant 6TG-resistant colonies would reflect the number of 6TG resistant cells in the stock culture. At each time point, a 1000 cells from the A2-4 cell line were plated in medium containing 6TG, as well as in non-selective medium, and the number of colonies that formed in 6TG medium was expressed as a percentage of those that grew on non-selective medium. This controlled for variation in the plating efficiency. In the cases of the M13 and 4A5 cell lines, 10^6 cells were plated in both media. In all cases each subculture was set up in duplicate. The mean number of 6TG-resistant colonies was calculated and plotted against the number of generations that had passed whilst on non-selective medium. It was assumed that the rate of reversion to 6TG resistance would remain constant at each generation and so the proportion of revertant cells would increase with a linear relationship to generation time. The UNIGRAPH package was used to fit a linear regression line to each set of data and the rate of reversion
Figure 4.1  The reversion rates of the cell lines A2-4, M13 and 4a5

The rate of loss of hpnt expression of each cell line was measured as the rate of reversion to 6TG resistance, see Section 4.2.

The regression lines were plotted using the UNIGRAPH graphics package running on a VAX 11/750.

A: A2-4 cells.
B: M13 and 4a5 cells.
Table 4.1 The frequency of reversion to 6TG resistance after transfection by calcium phosphate precipitation

<table>
<thead>
<tr>
<th>Cell line</th>
<th>Precipitate</th>
<th>Selection</th>
<th>Freq' of colonies</th>
</tr>
</thead>
<tbody>
<tr>
<td>A2-4</td>
<td>no</td>
<td>6TG</td>
<td>3.5 x 10^-3</td>
</tr>
<tr>
<td></td>
<td>yes</td>
<td>6TG</td>
<td>4.0 x 10^-3</td>
</tr>
<tr>
<td></td>
<td>yes</td>
<td>6TG &amp; G418</td>
<td>1.0 x 10^-6</td>
</tr>
<tr>
<td>M13</td>
<td>no</td>
<td>6TG</td>
<td>6.3 x 10^-5</td>
</tr>
<tr>
<td></td>
<td>yes</td>
<td>6TG</td>
<td>6.6 x 10^-5</td>
</tr>
<tr>
<td></td>
<td>yes</td>
<td>6TG &amp; G418</td>
<td>0</td>
</tr>
</tbody>
</table>
A: The map of p(Bam)neo, showing Clal, EcoRI, BamHI, EagI and PstI sites. The bacterial plasmid vector sequences are shown as a single line, inserts (neo coding sequence and SV40 early polyadenylation signals) as blocks. The construction of the plasmid is described in Section 3.2.

B: Proposed integration of p(Bam)neo into the genomic DNA

B: The possible modes of integration of the promoterless neo gene contained within p(Bam)neo to aligne the neo coding region with a genomic region (P) which expresses it. (Genomic DNA regions are represented as dotted lines.)

(i) Integration may occur at a point within the plasmid vector, therefore BamHI and EcoRI/BmaHI digestion will produce neo-hybridising fragments of the same size.

(ii) Integration may occur at a point within the 5' region of the neo coding region, therefore BamHI and EcoRI/BmaHI digestion will produce novel neo-hybridising fragments.
a)\[\begin{align*}
P_{\text{EagI}} (0.1) & \quad P_{\text{BamHI}} (1.2) \\
P_{\text{PstI}} (4.0) & \quad P_{\text{EagI}} (2.8) \\
P_{\text{BamHI}} (2.2) &
\end{align*}\]

b)\[\begin{align*}
(i) & \quad P_{\text{EagI}}/\text{EcoRI}/\text{BamHI} \\
(ii) & \quad P_{\text{BamHI}}/\text{EcoRI}/\text{BamHI}
\end{align*}\]
Table 4.2 Transfection frequencies with promoterless the plasmid, \( p(Bam)\)neo

<table>
<thead>
<tr>
<th>Cell line</th>
<th>Plasmid</th>
<th>Transfection Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>A9</td>
<td>pML2dMmtneo</td>
<td>(1.0 \times 10^{-4})</td>
</tr>
<tr>
<td>A2-4</td>
<td>pML2dMmtneo</td>
<td>(0.5 \times 10^{-4})</td>
</tr>
<tr>
<td>A9</td>
<td>(p(Bam))neo</td>
<td>(0.6 \times 10^{-6})</td>
</tr>
<tr>
<td>A2-4</td>
<td>(p(Bam))neo</td>
<td>(1.0 \times 10^{-6})</td>
</tr>
<tr>
<td>A9</td>
<td>(p(Bam))neo</td>
<td>(3.5 \times 10^{-6})</td>
</tr>
</tbody>
</table>

*linearised*
was estimated from the slope of this line (see figure 4.1). This gave the following reversion rates:

- A2-4: $0.78 \times 10^{-2}$ revertants per cell generation;
- M13: $1.50 \times 10^{-5}$ revertants per cell generation;
- 4A5: $0.29 \times 10^{-6}$ revertants per cell generation.

In a second experiment, the effect of calcium phosphate precipitation on the reversion frequency of the lines, A2-4 and M13, was investigated using the plasmid pML2dMmtneo to transflect flasks of $10^6$ cells from either A2-4 or M13 line. These were then put onto medium containing either 6TG or both 6TG and G418. Control cultures were grown of cells on non-selective medium for 3 days before transferring them to medium containing 6TG, which simulates the length of time cells were on non-selective medium during the transfection procedure. The results are shown in Table 4.1.

Although the results of this experiment show a slight increase of the reversion frequency following transfection of both of these cell lines, the increases were not significantly different from the reversion frequency without transfection (A2-4 $t=0.333$, 2 degrees of freedom; M13 $t=0.345$, 1 degree of freedom). On this evidence, transfection using calcium phosphate precipitation would not be expected to have a major influence on the reversion frequency.

### 4.3. Transfection of the plasmid p(Bam)neo into A9 cells

The second element which could contribute to the "background" is the frequency at which G418 resistance is obtained after
transfection with a promoterless neo gene. This was investigated by carrying out transfections using the plasmid p(Bam)neo, that was a plasmid containing the neo coding sequence from which Mint promoter had been removed, (see Section 3.2). This was transfected in two forms, as a covalently closed circular molecule or linear molecule formed by digestion with EcoRI, which cuts the plasmid at the 5' end of the neo gene. It is possible that a double strand break would result in the plasmid preferentially integrating at the point of the break. This may increase the probability that after integration the 5' region of the neo gene would be aligned with a region of cellular DNA, that is capable of promoting its expression (see Figure 4.2).

10^6 A9 and A2-4 cells per flask were transfected and selected on G418, with the plasmid pML2dMintneo being transfected as a control. The results, Table 4.2, show that a promoterless neo gene is expressed in both cell lines at a frequency of about 1 in 10^-6, a hundredfold less than achieved with an attached promoter. Linearising the plasmid appeared to increase the frequency slightly, but with consideration to the size of the sample, this difference was not statistically significant (t=1.06, 3 degrees of freedom).

There are a number of ways in which G418-resistant colonies could have been generated:

i) they may have been cells which spontaneously became resistant to G418, although resistant cells were never seen on controls at this concentration of G418;
Figure 4.3  Analysis of cell line generated after transfection with p(Bam)neo

A: 10 ug of DNA from the lines B11, B12 and B22 was digested with EcoRI or BamHI, separated by electrophoresis, blotted onto Biotrace-RP membrane, hybridised to $^{32}$P labelled p(Bam)neo and autoradiographed for 5 days.

A mixture of EcoR1 and BmaHI digested p(Bam)neo DNA was included to the equivalent to 100, 10 and 1 copy of the plasmid per cell.

The band of approximately 4.7 Kb in all the cell lines is likely to be due to a contaminating DNA which hybridises to the plasmid probe.

B: 10 ug of DNA from the lines B11, B12 and B22 was digested with EcoRI, BamHI or BamHI and EcoRI (B/E), separated by electrophoresis, blotted onto Biotrace–RP membrane, hybridised to a $^{32}$P labelled neo-specific probe and autoradiographed for 5 days.

The sizes shown are in Kb.

For experimental details see Section 4.3.
Probe: p(Bam)neo
ii) it was possible that the transfection was contaminated with the plasmid pML2dMMmtneo, so these colonies contain at least one copy of this plasmid;

iii) a low level expression may have occurred even in the absence of the promoter, such that if enough copies of the plasmid are present sufficient neo protein could be synthesised to confer G418 resistance;

iv) the 5' end of the neo gene might have integrated adjacent to a sequence that allows it to be expressed.

In the latter case, such a sequence could be a promoter, or an enhancer that could increase the level of expression to that which could give rise to a resistant cell. This sequence could either be of cellular origin or perhaps a cryptic site produced by rearrangement within the plasmid.

In order to investigate some of these possibilities, cell lines were established from three of these colonies, (B11, B12 & B22) and DNA was isolated from them. From the map of p(Bam)neo in Figure 4.2a, it can be seen that BamHI cuts the plasmid twice, excising a 2.2 Kb neo fragment, whilst EcoRI cuts the plasmid once, at the 5' junction between the plasmid and neo sequences. Integration into the chromosomal DNA may or may not be expected to interfere with the size of the 2.2 Kb neo containing BamHI fragment, as shown in Figure 4.2b. Figure 4.3a shows the hybridisation pattern of BamHI or EcoRI digests of the DNA to the p(Bam)neo probe. It is clear that plasmid is present in the DNA which argues that transfection of the neo gene is responsible for G418 resistance, eliminating option (i) above. From
the intensities of the bands, it is likely that there are one or two copies of plasmid in lines B11 and B12, whilst there are greater than ten copies in line B22. Overall, this argues against option (iii) above. In each case a 2.2 Kb BamHI fragment is present indicating that in each line some of the plasmid integrated was p(Bam)neo. In order to check that the transfection was not contaminated with the plasmid pML2dMmtneo, a double digest of BamHI and EcoRI was carried out on each DNA. These digests were separated by agarose gel electrophoresis, blotted and hybridised with a probe specific to the neo coding sequence. As the EcoRI site is situated adjacent to the 5' BamHI site of the neo gene in p(Bam)neo, the neo fragment after the double digest would be the same size as the one produced by a BamHI digest. A BamHI and EcoRI double digest of the plasmid pML2dMmtneo would produce a 4.1 kb fragment. The results of this digest were probed with a neo specific probe, and are shown in Figure 4.2b. In no case is there a 4.1 Kb band, which argues against option (ii) above. It is interesting to note that in each cell line there is a novel fragment that hybridises to the neo probe. These could result from the neo gene integrating into the cellular DNA through the 5' non-coding region of the neo gene and so becoming aligned with an endogenous promoter region. Hybridisation with the whole plasmid showed that in all cases there were rearrangements of the plasmids, and so it is also possible that the neo fragments were produced by rearrangements of the plasmids. In conclusion, G418-resistant colonies would appear to result from the expression of p(Bam)neo as a result of integration, but further analysis would be necessary to show what processes have resulted in the expression of the neo gene.
4.4. Discussion

Spontaneous loss of hprrt and functional integration of a promoterless neo gene have measurable frequencies, but at what frequency would a colony arise in which both of these events had occurred? In the case of the A2-4 cells, revertants arise at a rate of around 1% of cells per generation. During a transfection the cells are grown on non-selective medium for three days, therefore approximately 3% of the cell population would be expected to be 6TG-resistant. Assuming that the transfection frequency of $1 \times 10^{-6}$ cells for expression of the plasmid p(Bam)neo is representative of other promoterless neo genes, cells that are both 6TG- and G418-resistant would be expected to arise at a frequency of approximately $3 \times 10^{-8}$. Similarly, the M13 cell line would be expected to produce colonies that are both 6TG- and G418-resistant at a frequency of around $4.5 \times 10^{-11}$ cells.

Would these frequencies be expected to interfere with my scheme to isolate cells in which the integration of a plasmid has been targeted to the hprrt gene? As discussed in Section 1.9, 1/1000 of the transfectants would be expected to show targeted integration into the hprrt gene. If a transfection frequency of $1 \times 10^{-4}$ cells is achieved, then $1 \times 10^{-7}$ cells would be expected to show targeted integration. For the M13 cell line, the background of $4.5 \times 10^{-11}$ is ten thousandfold less than the expected frequency of targeted integration and so would not be expected to be a problem. For the A2-4 cells, however, the background is only tenfold less. This therefore may cause problems in the targeted integration experiments. Despite the potential problems, it was decided to use the A2-4 cells as they possess the high transfection frequency that is needed to obtain any colonies at all. In addition, as the majority of the genome is derived from a mouse.
cell, there are no human hprt pseudogenes present, this would make the analysis of the integration sites by hybridisation simpler.

The stability of the hprt genes in the cell lines investigated follow the pattern expected, although there are some interesting points. As expected for the A2-4 cell line, the human X-chromosome is lost at a high frequency, but it is also interesting to note that both the transfectant lines revert to 6TG resistance at a higher frequency than would be expected. For example, hprt\(^-\) L cells have been reported to arise spontaneously at an average rate of \(5 \times 10^{-7}\) cells per generation (Littlefield, 1964). This difference in reversion frequency of "stable" clones has been noted in earlier work. For example, two stable murine cell lines, made hprt\(^+\) by transfection with total cellular DNA, were found to revert to 6TG resistance at \(8 \times 10^{-5}\) and \(4 \times 10^{-4}\) cells per generation (Graf et al, 1979). These figures compared with the control figure of less than \(6 \times 10^{-8}\) cells per generation. It has been shown that in transfectant cell lines expression is lost from integrated plasmids with a frequency that varies between \(1 \times 10^{-3}\) to \(1 \times 10^{-5}\) cells per generation. There are a number of ways in which the expression may be lost, such as deletion, methylation and amplification (Gebara et al, 1987).

It is interesting to note that the cell line 4A5, which has two separate integrations of the plasmid p4A8, reverted at a frequency only tenfold less than that of the M13 line and of those reported. Although the fact that it was reduced with respect to the M13 reversion frequency is consistent with the argument that two events are required to lose both genes, it is not clear why the frequency is only tenfold less.

It was surprising that a plasmid containing a promoterless neo gene produced G418-resistant colonies at a frequency only a
hundredfold less than a plasmid containing a neo gene expressed from a promoter, although it is consistent with the transfection frequencies of other plasmids containing promoterless neo genes (Colbere-Garapin et al, 1981; Southern & Berg, 1982). It shows that in around 1% of the sites where a plasmid integrates, a genomic sequence capable of actively expressing a gene becomes linked to the neo coding sequence. This is much higher than might be expected intuitively. This suggests that integration may occur preferentially into transcriptionally active chromatin, as suggested with retrovirus integration (Section 1.5).

If this is the case, it may be possible to use this as a method to randomly isolate novel cellular promoters. A similar approach has been used by Hamada (1986) to isolate regions of cellular DNA that enhance the expression of an enhancerless pSVgpt construct.

In conclusion, a background of cells which have failed to target the plasmid to the hprt gene must be expected at the frequencies given above. This background would not be expected to be a problem when attempting to target integration into the hprt gene of the line M13, but it may become a problem when using the A2-4 cells.
CHAPTER 5

CONSTRUCTION OF THE HPRT/NEO FUSION GENE
5.1. Introduction

In the Introduction, a strategy for the biochemical selection of those cells which have integrated a plasmid by homologous recombination was outlined. This requires a plasmid that contains a gene which, in addition to a region of sequence homology to the hprt chromosomal locus, can be selected as a neo gene. In this chapter is described the construction of an hprt/neo fusion gene that has these properties.

The fusion gene must be designed so that, after integration into the chromosome, it is expressed from the hprt chromosomal promoter to produce a protein with the activity of the neo gene. For this to occur by a process of homologous recombination the hprt sequence must precede the neo gene and be in the same orientation. In principle, there are two ways of constructing such a gene:

i) make a construct which produces a bicistronic mRNA molecule, i.e. a mRNA that contains separate coding sequences of both hprt and neo, so that translation of the hprt coding sequence is followed by re-initiation and translation of the neo coding sequence;

ii) make a gene which contains a single coding sequence that is a fusion of both hprt and neo components.

Peabody and Berg (1986) have made recombinant plasmids which express bicistronic genes in the manner proposed in (i) above. As I was unaware of any evidence that this may occur in mammalian cells when I started the project, I chose to construct the second form of
fusion gene. In this chapter I will use the term "fusion gene" to describe this second form of construction.

5.2. Principles of the design of the hprt/neo fusion gene

DNA sequences from a number of sources have been fused to the neo coding sequence, and expressed in *E. coli* to produce proteins in which varying numbers of amino acid residues were fused to either the N or the C terminus of the neo polypeptide (Reiss et al, 1984). All fusion genes were capable of conferring kanamycin resistance, however there was a large variation in the level of the resistance achieved. As the number of amino acids residues at N-terminus of the fusion protein was increased, the level of resistance decreased. When the number of residues was greater than 100, the proportional relationship of length to level of resistance no longer held. Some large additions were much more active than smaller ones, eg. 304 residues from the penP gene of *B. licheniformes* resulted in a fusion protein that was more active than those formed with smaller additions from the same gene. This was explained by suggesting that additions of long polypeptides to the N-terminus generally caused large reductions in the activity of the neo component, but that in some cases a higher than expected level of activity resulted from proteolytic cleavage of the neo component away from the added polypeptide. As the ability to cleave the fusion gene depends on the protein structure, the activity is not related to the length of the additional polypeptide (Reiss et al, 1984).

In order to maximise the length of homology between the chromosomal site and plasmid, it was decided to construct a fusion gene that contains 480 bp of hprt sequence. If the polypeptide proved to be inactive, the size of the hprt sequence would be systematically reduced until a functional gene product was obtained.
Figure 5.1 The general scheme for the construction of the hprt.neo fusion gene

This illustrates the scheme to insert a linker into the neo gene of pML2Mmtneo, disrupting the expression of the gene by introducing a frame-shift mutation (marked by a cross) to make pLink. Appropriate restriction enzyme sites would then be introduced into the human hprt cDNA so that it may be ligated into the disrupted neo gene and restore the correct reading frame.
Insert KpnI linker to disrupt the neo gene

Introduce KpnI sites into hprt cDNA

Ligate hprt KpnI fragment into pLink
The expression of the neo gene from the Mmt promoter in both E. coli and mammalian cells (see Chapter 3), suggested that, initially, it would be possible to test the fusion gene in E. coli before introducing it into mammalian cells. This would enable rapid testing and construction of the fusion gene.

In bacteria, the 5' non-coding region of a gene is important for efficient translation. In order to get efficient expression of the fusion gene in E. coli, it was important to retain this region of the neo gene after the addition of the hppt DNA. The hppt sequence was therefore inserted into the protein coding sequence of the neo gene. Unfortunately neither the hppt nor the neo gene contained convenient restriction enzyme sites for this insertion and so linkers were needed to generate both the insertion site in the neo gene and the compatible ends on the hppt DNA. It was reasoned that the insertion of a linker into the neo gene could have a second use. The presence of the linker could introduce a frame-shift mutation disrupting the neo coding sequence gene, such that after ligation of a suitable hppt fragment into the linker, the original reading frame would be restored. It should be possible to identify the functional fusion gene as it would confer kanamycin resistance on E. coli.

In this chapter is described the construction of a fusion gene using these principles. Firstly, the genetic manipulations of the hppt and neo sequences were modelled in the computer (Section 5.3) Then these and additional manipulations were carried out in order create the fusion gene (Sections 5.4-5.6) and some of its properties were studied (Section 5.7). In Figure 5.1, the major genetic manipulations needed to construct the fusion gene are illustrated.
5.3. Computer modelling

Before attempting to construct the fusion gene the following questions needed to be answered:

i) is there a suitable site in which to insert the linker?

ii) what linker sequence would both create a novel restriction enzyme site and cause a frame-shift mutation?

iii) what fragment of the hprt gene would compensate for the frame-shift mutation caused by insertion of the linker?

The programs FETCH, SEQED and MAP from the Wisconsin (UWGCG) sequence analysis package were used to solve these problems.

The coding sequence of the neo protein was examined using MAP to predict potential restriction enzyme sites. An unique EagI site was found 35 bp 3' of the neo initiation codon and was an ideal position to insert the linker. SEQED was used to make additions to the neo sequence at the EagI site in the computer that would be equivalent to the effects of introducing a linker into the site, and then MAP was used to predict the consequence of each addition on the translation of the neo gene. In this way it was possible to arrive at a suitable linker sequence by trial and error. The oligonucleotide, GGCCTGGTACC, which by self annealing forms a complete KpnI site with EagI sticky ends (see Figure 5.2a) was found to be satisfactory as it causes a frame-shift mutation that would be expected to result in the premature termination of the neo protein after 25 residues. This was designated neo.link.

The computer was then used to predict which hprt fragments could be used to correct the frame-shift mutation. As the hprt cDNA did not
Figure 5.2  Computer modelling of the construction of the hprt.neo fusion gene

A: Insertion of a EagI/KpnI linker creates a frame-shift mutation which leads to premature termination of translation of the neo gene.

B: Creation of a novel hprt KpnI fragment by oligonucleotide site directed mutagenesis and by introduction of a KpnI linker.

C: Insertion of the hprt KpnI fragment creates a fusion gene with the correct reading frame of the hprt and neo genes (shown by the dotted line and box respectively). Insertion of the hprt fragment in the opposite orientation is shown to result in termination of translation occurs within the inserted region.

The nucleotide numbering of the hprt and neo sequences corresponds to that of the Genbank database. The nucleotide numbering of the fusion gene is a result of the default of SEQED.
hprt

Oligonucleotide site-directed mutagenesis

Cut with HindIII, add the linker CGGTACCC

ThrArgSerProGlyValValLeuSerAsp

TyrAsnProLysMetValLysValAlaSerLeuLeuValLysArgThrProArg
contain any KpnI sites, it was necessary to take into account how the KpnI sites could be added. There are two ways in which to introduce sites into the hprt sequence, either by ligation of a KpnI linker into a suitable site or by site directed mutagenesis.

Figure 5.2b shows the regions of hprt sequence where the KpnI sites were introduced. The sequence starting at position 89 is GGGTACC, this differs from the KpnI site, GGTACC, by only two nucleotides, shown in bold. It was therefore possible to attempt to alter this sequence to that of a KpnI site by oligonucleotide site directed mutagenesis. It was also possible to find similar sequences that could have been used to generate the 3' KpnI site of the hprt fragment. As greater flexibility was desired with the position of the 3' KpnI site in case the original design proved not to be active, it was decided that a linker could be both ligated into a predetermined site and to fragments generated randomly by limited exonuclease digestion. The KpnI linker, CGGTACCG, was inserted into the HindIII site, at position 568. The introduction of both KpnI sites into hprt is illustrated in Figure 5.2b. The outcome of ligating the final hprt KpnI fragments into neo.link was tested using the SEQED and MAP programs, and as shown in Figure 5.2c, the reading frame required to produce the neo protein is restored.

Figure 5.2c also shows that if the hprt KpnI fragment is ligated to neo.link in the opposite orientation, then premature termination will occur in the hprt sequence after 73 residues. It was predicted therefore that it should be possible to distinguish between the two orientations on the grounds that only the one orientation will produce a gene that can confer kanamycin resistance on E. coli.
pML2dMmtneo

Subclone into pGEM

pGEMMmtneo

Remove KpnI site:

pGEMMmtneo (KpnI)

Insert KpnI linker into EagI site

pGEMlink

Subclone

Mmt-neo.link to pML2d
Figure 5.4  Analysis of pLink

Bacterial colonies pGEMlink 3, 19 and 40 were found to have a KpnI site. The insertion of this into the EagI site was confirmed by a BglII/KpnI digest. This was expected to produce a unique 70 bp fragment.

The figure shows the results of the end-labelled products of such digests separated on a 8% polyacrylamide gel and autoradiographed for 1 day. The 70 bp fragment is marked (*).

To control for complete digestion Lambda HindIII marker DNA was included in the digests. The 60 bp fragment expected from complete digestion of the Lambda DNA is marked (A). The 125 bp Lambda HindIII fragment is also marked (A).

Track M contains pGEM2 digested with HinfI and end-labelled, track M' contains M13mp8 digested with HinfI and end-labelled. The sizes shown are in bp.
5.4. Insertion of a linker into the neo gene

The computer modelling was based only on the neo sequence. Unfortunately when the neo gene was considered as a part of the plasmid pML2dMmtneo, there was an additional EagI site in the pML2 component of the plasmid and a KpnI site in the Mmt promoter which made the proposed manipulations more complex. In order to carry out the construction successfully it was necessary to remove both of these sites. As illustrated in Figure 5.3, this was done in two separate manipulations. Firstly, the EcoRl/BamHI fragment of pML2dMmtneo, containing the Mmt-neo sequence, was subcloned into the plasmid pGEM2, which contains no EagI site, to produce pGEMMmtneo. It was then possible to remove the KpnI site in the Mmt promoter by linearising pGEMMmtneo with KpnI, removing the single stranded ends using T4 DNA polymerase, recircularising the plasmid using T4 DNA ligase and the product used to transform E. coli HB101 cells. The resulting plasmid, which had no KpnI sites, was then linearised with EagI and ligated to the KpnI/EagI linker, which had been made and purified on site. This was used to transform E. coli DH5 cells which were then selected on ampicillin plates. Kanamycin sensitive clones were identified by replica plating onto ampicillin and kanamycin plates. Plasmids derived from three kanamycin sensitive colonies were analysed by restriction enzyme digestion. All three were found to contain a single KpnI site, that mapped to the position expected. As a further test that the fragment was inserted into the EaqI, the plasmids were digested with BglII and KpnI, end-labelled and separated on an 8% polyacrylamide gel. If the KpnI site had been ligated into the EagI site it would have been expected to create a fragment of 70 bp in size after digestion with BglII and KpnI. This was shown to be the case (see Figure 5.4). Finally the EcoRl/BamHI fragment of the plasmid was
Table 5.1 The distribution of the level of resistance to kanamycin of individual colonies that contain the plasmid pLink

<table>
<thead>
<tr>
<th>Percentage of colonies(^{(a)})</th>
<th>Level of kanamycin resistance(^{(b)}) (ug/ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>30</td>
</tr>
<tr>
<td>2.5</td>
<td>50</td>
</tr>
<tr>
<td>40</td>
<td>75</td>
</tr>
<tr>
<td>17.5</td>
<td>100</td>
</tr>
<tr>
<td>17.5</td>
<td>150</td>
</tr>
<tr>
<td>0</td>
<td>200</td>
</tr>
</tbody>
</table>

(a) Expressed as a percentage of the original 40 colonies picked.

(b) Assessed by replica plating the original 40 colonies onto a set of plates containing concentrations of kanamycin that varies from 30 to 200 ug/ml of kanamycin (see text).
Figure 5.5  The manipulations required to create an hprt KpnI fragment

This figure illustrates the manipulations required to make the M13 phage (T2) that contains an hprt fragment with the two KpnI sites. The construction of M13mp8F and 8HiPs is described in Section 3.5, and the other constructions in Section 5.5.
Subclone PstI/BamHI fragment into M13mp8

Subclone PstI/HindIII fragment into M13mp8

Introduce KpnI site by oligonucleotide site directed mutagenesis

Introduce KpnI linker into HindIII site

Subclone into M13mp19

Ligate XhoI/BglII fragments

T2
recloned into the plasmid pML2d to generate the plasmid called pLink.

Figure 5.11 shows an interesting feature of pLink. Although it did not confer resistance to kanamycin at concentrations of 50 μg/ml, it did confer resistance on plates containing up to 30 μg/ml kanamycin. It was decided to test whether these colonies arose by reversion of pLink to pML2dMmtneo through excision of the linker by picking colonies grown for 48 hours on a plate containing 30 μg/ml of kanamycin and replica plating them onto plates containing kanamycin at concentrations ranging from 0 to 200 μg/ml. The results are shown in Table 5.1. A wide range of maximum resistances were scored, but no colonies were found to grow at the same rate as the controls containing authentic pML2dMmtneo. This argues against loss of linker and suggests that the frame-shift mutation must not completely abolish translation to produce a functional neo protein. It would appear therefore to be a potential example of "translational frame-shifting". The difference in resistances must be due to variation between individual cells. As pLink was not found to produce colonies on a 50 μg/ml kanamycin plate overnight, this phenomenon did not interfere with the selection for the fusion gene.

5.5. Creating the hprt fragment

The 5' and 3' KpnI sites were introduced into hprt sequences contained in separate M13 phage subclones that had been generated during sequencing (see Section 3.5 & Figure 3.7). This is illustrated in Figure 5.5. Oligonucleotide site-directed mutagenesis was carried out on the single-stranded DNA form of the M13mp8F construct, using the method described by Zoller and Smith (1982). Although this is the simplest method, it is not very efficient as the mismatched duplex is repaired at a high frequency after transformation. It is necessary
therefore to screen by hybridisation with the synthetic oligonucleotide for those phages that incorporate the required mutation. The Wallace rule relates melting temperature ($T_m$) of a duplex to length and base pair composition in the following way:

$$T_m = 4(\text{no. of G:C bp.}) + 2(\text{no. of A:T bp.})$$

Mismatched sequence results in a reduced $T_m$ with respect to the fully complementary sequence. This difference can be exploited to distinguish M13 phage containing the mutation from those that do not, since, after hybridisation, a labelled oligonucleotide, which perfectly matches the sequence will remain bound to a phage containing the mutation at a higher temperature than one that contains mismatches (as is the case with the oligonucleotide containing the mutation bound to the wildtype phage). Generally this is done by using as a probe the same oligonucleotide that was used to generate the mutant. The problem with using one oligonucleotide for both processes is that it must be capable of binding to the wildtype sequence under one set of conditions to introduce the mutation, and yet have a difference in $T_m$, following hybridisation to wildtype or mutant phage in order to differentially screen. This compromise generally results in only a small difference in $T_m$ between when it is bound to the mutant and the wildtype. In practice, when the mutation generates only a small change in the $T_m$, the difference in signal due to the presence of the mutation may be less than the difference due to variation in the DNA concentration, and this causes problems identifying true positives during screening. In order to improve the chances of positively identifying the desired mutants, I used two oligonucleotides. Mutagenesis was carried out using a 23-mer of the sequence
Both mutant and wildtype hprt sequences were sequenced using the dideoxy termination method and the products separated on the same gel. The five mutations are indicated. The KpnI site of the mutant is also indicated (++++++).
Mutant

A C G T

Wildtype

A C G T
GGCTCCGAGGTTGTTACCGGCAG, the changes in sequence are written in **bold**. This introduced three more mutations in addition to the two changes which generate the KpnI site. As these were 5' of the KpnI site they did not interfere with the final product. These additional changes also increased the G:C bp content of the mutant (two additions to the one G:C base pair lost in forming the KpnI site). A shorter 12 bp oligonucleotide, with the sequence **CGAGGTGGTAC**, was used for the screening. In this way, the predicted $T_m$ difference between mutant and wildtype sequence was increased to 15°C. This enabled the identification of two mutants out of the 70 plaques screened on the first attempt, and, in fact, the maximum difference in signal was obtained after washing the filter at 10°C below the $T_m$ predicted for probe annealed to the mutant sequence. These mutants were then checked in two ways:

i) by sequencing, Figure 5.6 shows the sequence of the M13/hprt mutant 463 compared with the wildtype hprt;

ii) by digestion of the RF form with KpnI.

The 3′ KpnI site was created by using the RF form of M13mp8FHips which contains only a single HindIII site, derived from the site within the hprt cDNA. The KpnI linker CGGTACCG was ligated into the HindIII site of the RF form of M13mp8FHips and the resulting construct used to transform the host JM101. Single stranded DNA was made from the resulting plaques and the dideoxy chain termination sequencing method was used to identify a phage that contained a single linker and no loss of nucleotides from the HindIII site. The subclone **link36** was found to be satisfactory.
The plasmid digests were separated on a 0.7% agarose gel, stained with ethidium bromide and photographed.

The gel was then blotted onto Biotrace-RP, hybridised to an hprt-specific probe and autoradiographed for 5 hours.

The following digests were carried out:

- a, pAH-1 with KpnI
- b, pAH-1 with XhoI
- c, pAH-1 with BamHI and BglII
- d, pAH-1 with BamHI, BglII and XhoI
- e, pAH-2 with KpnI
- f, pAH-2 with XhoI
- g, pAH-2 with BamHI and BglII
- h, pAH-2 with BamHI, BglII and XhoI
- i, pLink with BamHI and BglII
- j, pLink with KpnI
- k, pLink undigested
- l, pLink and pGEMhprt with XhoI
- m, pGEMhprt with XhoI

From the digests and patterns of hybridisation, it can be seen that pAH-1 and pAH-2 are identical and contain an hprt-hybridising fragment of approximately 500 bp and a XhoI site that was not present in pLink. The BamHI/BglII digests of pAH and pLink show that the neo fragment contains the insert. The hybridisation to pGEMhprt demonstrates that only hprt sequences were detected. This is consistent with the map of pAH (C).

The sizes shown are in Kb.
Probe: hprt
The two KpnI sites at this point were in hppt fragments contained in different M13 subclones. It was possible to bring the two sites together by ligation through the unique XhoI site common to the hppt sequences of both subclones. In order to do this it was necessary to use the unique BglII site within the M13 component of each construct, but unfortunately the hppt sequence of the mutant 643 was in the opposite orientation to that of link36. The PstI/BamHI hppt fragment of 643 was subcloned into M13mp19, thereby reversing the orientation. The small XhoI/BglII fragment of M13mp19-463 was ligated into the large XhoI/BglII fragment of link36 to produce T2, which contains both KpnI sites.

5.6. Inserting the hppt fragment into pLink

The KpnI fragment of T2, which contains the hppt sequence, was ligated into the phosphatase-treated KpnI site of pLink and the ligation mix used to transform the host DH5. Some of the transformed cells were plated onto a kanamycin plate (50 ug/ml). Figure 5.7a shows the digests of plasmids derived from two resistant colonies, termed pAH-1 and pAH-2. From the figure it can be seen that both of the plasmids are approximately 500 bp larger than pML2dMmtneo and have gained a XhoI site. This was found to be the case for all colonies picked from the kanamycin plate. By hybridisation using an hppt-specific probe, it can be seen that only the pAH plasmids contain hppt sequences (Figure 5.7b). KpnI digestion produced a band that migrates with the same mobility as the single cut pML2dMmtneo plasmid. When this digest was analysed by hybridisation to the hppt-specific probe, the hppt sequence is localised to a band with a size just less than 500bp, which could not easily be seen in ethidium bromide-stained gels. This is consistent with the expected map of the fusion gene
Figure 5.8  Analysis of the orientation of the hprt insert of pAH and pREV27

Plasmid DNA of pAH-1, pAH-2 and pREV27 was digested with XhoI, end-labelled and then digested with BglII. The products were separated on a 6% polyacrylamide gel and autoradiographed. Both pAH plasmids contained the 220 bp fragment expected for the functional fusion gene, whilst pREV27 contained the 460 bp fragment expected for the inactive reverse orientation.

The tracks marked M contain an end-labelled pGEM2 HinfI digest. The sizes shown are in bp.
Figure 5.9  Removal of the Mmt promoter from pAH

A: The plasmid digests were separated on a 0.7% agarose gel, stained with ethidium bromide and photographed.

B: The gel was then blotted onto Biotrace-RP, hybridised to a Mmt-specific probe and autoradiographed for 5 hours.

C: The map of pAH(M−).

The following digests were carried out:

a, pAH with PstI and KpnI  
b, pAH with PstI  
c, pAH with EcoRI  
d, pAH with BamHI  
e, pAH with XhoI  
f, pUC19 with EcoRI  
g, pAH(M−) with PstI and KpnI  
h, pAH(M−) with PstI  
i, pAH(M−) and pUC19 with EcoRI  
j, pAH(M−) with BamHI  
k, pAH(M−) with XhoI (partial)  
j, pAH(M−) undigested.

The results of the digest are consistent with the map (C) and hybridisation shows the absence of Mmt fragments in pAH(M−).

The tracks marked M contain Lambda DNA digested with HindIII. The sizes shown are in Kb.
pAH (M)
construct (Figure 5.7c). It was concluded that the scheme as modelled on the computer had resulted in a plasmid which contains an hprt/neo gene that confers kanamycin resistance to bacterial cells.

The computer model predicted that only one orientation would permit translation of the neo coding sequence and so only one orientation was expected to give kanamycin-resistant colonies. The remainder of the transformed bacterial cells were plated onto an ampicillin plate and colonies, which contained hprr sequences but were kanamycin-sensitive, were isolated by colony hybridisation to an hprr-specific probe and by replica plating colonies onto kanamycin plates. A plasmid DNA from a colony isolated in this way, pREV27, and the plasmids pAH-1 and pAH-2 were analysed for the orientation of the insert. Each plasmid was cut with XhoI, end-labelled and further digested with BglII before separation on a 6% polyacrylamide gel. The orientation that was predicted to result in a functional neo gene would be expected to have a 220 bp BglII/XhoI fragment, whilst a gene with the reverse orientation would be expected to have a 460 bp fragment. Figure 5.8 shows the autoradiograph of this gel and the results are consistent with the predictions of the computer model.

The plasmid pAH-2 was chosen as the basis for the homologous recombination experiments. The Mmt promoter was removed by digestion with EcoRI and BglII, the ends filled-in and the plasmid recircularised. The removal of the promoter was confirmed by gel electrophoresis and hybridisation to a Mmt-specific probe, (see Figure 5.9). This construct was named pAH(M-).

5.7. Testing the kanamycin resistance of the constructs

In view of the results of Reiss et al (1984), it was of interest to see what level of kanamycin resistance would be conferred by the
Figure 5.10  The maximum level of resistance conferred by pLink, pML2dMmtneo, pAH, pREV27 and pAH(M−)

This figure shows a histogram of the maximum level of kanamycin resistance of E. coli DH5 and E. coli HB101 after transformation with pLink, pML2dMmtneo, pAH, pREV27 and pAH(M−). This was measured using the method of Reiss et al (1984), and is described in the text (see Section 5.7).
Figure 5.11  Identification of putative promoters within the hpri
insert of pAH

This figure shows the sequences of the putative bacterial gene promoters indentified using the Find program of the UWGCG sequence analysis package, and their respective homologies to the consensus promoter. The nucleotides in common with the consensus are boxed.

In addition to the putative promoters, that of the promoter for the tetracycline resistance gene from pBR322 and those identified in Section 3.3 are shown.
CONSENSUS SEQUENCE

TTGACAT --------------------- TATAAT
-35
-10

HOMOLOGY SCORE

66%

PROMOTER

-35
-10

Mt PROMOTER

GCAACCTTCTGCCCCTTTCTCTGATAGATAGAG
-35
-10

HOMOLOGY SCORE

41%

PROMOTER

-35
-10

Cryptic (i)

TCCCCGAAATTCTGTTAACATATATATCAT
-35
-29
-10

Cryptic (ii)

(T) CTAAGAACCATTATA
-35
-31
-10

Cryptic (iii)

(G) TGAAGATATATGAA---
-35
-31
-21
-10

HOMOLOGY SCORE

i) 32%

ii) 49%

iii) 59%

hprt INSERT

---ATGCAGACTCTTTGGTCAGGCAGAT
cAAA
-35
-21
-10

HOMOLOGY SCORE

49%

REVERSE ORIENTATION

TGCATTTTTGCAGATATCTTT
-35
-28
-10

HOMOLOGY SCORE

41%

TAGCCCGCCCTGGAGCAAGAGGCTCACATGTG
-35
-28
-10

HOMOLOGY SCORE

50%
The hprt/neo fusion gene constructs. Figure 5.10 shows the maximum concentrations of kanamycin to which each plasmid conferred resistance. Surprisingly plasmids pAH and pAH(M-), both conferred a greater resistance than did pML2dMmntneo from which they were derived. Two possible explanations for this observation have been investigated:

i) the hprt insert may contain a bacterial promoter;

ii) the insertion of the hprt sequence results in an increased copy number of the plasmid and therefore the total amount of fusion protein produced.

Figure 5.11 shows the results of a search for potential bacterial promoters using the program FIND, as described in Section 3.3. Two hexanucleotides of sequence TATAAT, i.e. sequences with exact homology to the -10 element of the consensus bacterial promoter, were found. These were situated within the hprt insert 30 and 90 bp upstream from the 3′ junction with the neo sequence. These elements form four putative promoters which have a homology score of 50%, 48%, 59% and 49% (see Section 3.3). None however has the 17 bp spacing characteristic of the consensus promoter. The spacing of two promoters, marked (ii) and (iii) on Figure 5.11, appeared to be too small to function, whilst that of promoter (i), appeared too large. This leaves the putative promoter based around the second TATAAT element as the most likely candidate for the active promoter in the kanamycin resistance assays. This has a homology score of 49%.

For functional neo protein to be produced there must also be the bacterial signals for translation, which include an initiation codon in frame with the neo coding sequence and a 5′ sequence that contains a ribosome binding site. There are three possible initiation codons
that are in frame with the neo gene. The first is an AUG codon, but this is only 10 bp downstream of the second TATAAT element which may not be sufficient 5' sequence for efficient ribosome binding. The next possible initiation codon is 50 bp downstream of the second TATAAT element and is a GUG codon, which may act as an initiation codon in E. coli. If the gene is transcribed from the promoter with the second TATAAT element, again this may be too close to the 5' end of the mRNA to function efficiently. There is a third GUG initiation codon 100 bp downstream of the second TATAAT element which would allow enough 5' sequence for translation of the transcript from any of the promoters. Do any of these 5' sequences have similarity to those of efficiently translated genes? By using the W101 weight matrix, as in Section 3.3, each of the possible initiation codons was tested. They score, respectively, -210, -349 and -410, in comparison to a functional initiation codon which would be expected to score at least 2. The Mmt-neo gene initiation codon scores 17. This would argue that, although there are sequences with homology to the bacterial promoter in the hprt sequence and that these are present at a location suitable for expression of the neo component of the fusion gene, there is not enough sequence homology to the 5' regions of the bacterial gene to conclude that there would be efficient expression from these hypothetical promoters.

It is possible to find three putative promoters within the reverse orientation construct, they score 48%, 41% and 50%, but are situated at the 5' end or in the middle of the insert. Not only are they situated some distance from the hprt/neo junction, there are no available open reading frames that extend to the neo coding sequence.

Since the analysis of the hprt sequence did not reveal the presence of any likely bacterial promoters, an investigation was made
Figure 5.12 Analysis of the copy number of the pML2dMmtneo, pAH and pAH(M)

The experimental details are described in Section 5.7.

Undigested DNA from triplicate minipreps of the plasmids pML2dMmtneo, pAH and pAH(M) were made. 20 ul, 10 ul and 10 ul, respectively, were separated on a 0.7% agarose gel, stained with ethidium bromide and photographed.
into the copy number of the plasmids pML2dMmtneo, pAH and pAH(M-). Triplicate minipreps were made for each plasmid from 5 ml cultures standardised to a cell concentration giving 1 OD550 unit. One fifth and one tenth of the total miniprep DNAs from the pML2dMmtneo and the pAH/pAH(M-) plasmids, respectively, were loaded onto a gel (Figure 5.12). There was a large difference in the copy number of the plasmid pML2dMmtneo in comparison to pAH and pAH(M-). The amount of RNA at the bottom of the gel controlled for any large fluctuation in the preparations. A photographic negative of the gel was scanned and the relative amounts of DNA calculated. There was an approximate forty-fold difference in the amount of plasmid DNA between pML2dMmtneo and the pAH plasmids. By comparison with the intensities of the bands of the marker tracks, it was possible to calculate the amount of DNA and hence an approximate copy number for each plasmid. This suggested that there were only a few copies of pML2dMmtneo and around 100 copies of the pAH plasmids. To establish what may have caused the increase in copy number, the investigation was extended to cover pLink and pREV27. Both of these also had a copy number equivalent to that of pAH. The increase in copy number appears to have occurred with the generation of pLink. This could have been a result of the insertion of the linker into the neo gene or a secondary mutation of the plasmid used as a basis for the later constructions.

5.8. Discussion

The aim of the work discussed in this chapter was to construct a gene which has homology to the human hprt gene but has the activity of the neo protein. The construction followed from a theoretical concept through computer-aided-design to produce the plasmid pAH. At the DNA level, the gene has a structure consistent with the computer
predictions and confers kanamycin resistance on E. coli. The question which was addressed in Section 5.7 was whether it actually produces a protein that is a fusion of polypeptides from hprt and neo. This is an important point because if the gene does not produce a fusion protein in the bacterium, it cannot be an adequate model for the function of the gene in a mammalian cell.

There are a number of possible explanations for the higher level of kanamycin resistance obtained following transformation with the fusion gene as compared with the original gene. This difference was investigated in two ways. Firstly, I searched for the inadvertent introduction of a bacterial promoter that could initiate expression of only the neo component of the fusion gene. Although a cluster of four potential promoters was found, detailed analysis of their sequence showed no evidence that they could express the neo coding sequence. These, however, are only predictions based on the observations of the bacterial promoters sequenced to date, and it may be that they are promoters which function but by mechanisms not yet taken into account by the analysis. Further work would be required to conclude definitely whether these potential promoter sequences do function. The most conclusive experiment would be to see if the mRNA produced by the fusion gene is approximately the same size as that transcribed from the original construct, or whether the fusion gene transcript contains an extra 500 bp of hprt sequence. Unfortunately, due to the difficulty in isolating bacterial mRNA, this is not an easy experiment to attempt. To look at the proteins would be less conclusive as the protein produced by the fusion gene would be expected to be cleaved into a neo and hprt component, and so a simple size change of the neo gene may not be diagnostic of production of the fused protein. The only effective way to investigate the phenomenon further is to attempt
to use the hprt fragment to express other marker genes, such as lacZ.

An alternative explanation for the increase in kanamycin resistance is that the increased copy number of the pAH plasmids results in the production of greater amounts of neo protein, and hence kanamycin resistance, than from pML2dMmtneo. The increase in overall expression of the fusion protein may mask the differences between the levels of expression from the cryptic and Mmt promoters that were identified in Section 3.3, and so result in a similar level of kanamycin resistance for both pAH and pAH(M'). It still remains a possibility, however, that with pAH at a high copy number, the putative promoter within the hprt insert may produce sufficient mRNA to confer kanamycin resistance despite weak translation initiation sequences.

The change in plasmid copy number has been traced to the point at which pLink was constructed. It is therefore independent of the presence of the hprt sequence. What has caused this increase in copy number? It may be that the neo gene in some way disrupted a sequence that normally maintains a low copy number. It is difficult to imagine how the neo gene is involved in copy number control. A second possibility is that a plasmid was isolated which carried an additional mutation in one of the genes known to control copy number. There are three genes in pBR322 known to control copy number: RNAI, RNAII and rop (reviewed by Davison, 1984). All copy number mutations identified so far have been localised to defects of the RNAI gene, the product of which has an inhibitory effect on the regulation of plasmid replication. For example, the very high copy number plasmid, pHCL, which arose spontaneously, was shown to be due to a single point mutation in the RNAI gene (Boros et al. 1984). I therefore suspect that pLink may also have a mutation in this gene. This is a useful attribute for pLink, because if it is used to construct other neo
fusion genes the increased copy number may compensate for the reduced activity which may result from the addition of the extra amino acid residues.

An additional observation made whilst constructing the fusion gene was that the frame-shift mutation of neo.link does not completely eliminate kanamycin resistance. The best explanation of this is that the ribosome undergoes a frame-shift as it translates the gene and so corrects the mutation. This has been documented in a number of cases and two groups of effects have been observed. Low levels of beta-galactosidase activity, around 0.1% of the wildtype, have been reported from frame-shift mutants of the LacZ gene and appear to be due to a mechanism of spontaneous frame-shifting (Atkins et al, 1972). Taking into account the increase in copy number with respect to the wildtype and the difference in the assay, spontaneous frame-shifting may account for the kanamycin resistance conferred on bacteria by the plasmid pLink. Specific amino acid codons, such as the alanine codon GCA and perhaps also GCT, have been found to be inherently more "shifty" (Bruce et al, 1986). There is an alanine residue with the sequence GCT situated four residues before the point at which premature termination would occur in neo.link and this could be a site at which frame-shifting may occur. Frame-shifting has also been observed to occur at a high frequency in the release factor 2 (RF2) gene of E. coli, which appears to frame-shift at a rate as high as 50% of all translations and may use frame-shifting as a mechanism of gene regulation (Craigen & Caskey, 1986). The basis for this control is not known. The neo.link gene does not appear to frame-shift as frequently as in the case of RF2.

The primary aim of producing a gene that contains both hprt sequence and an active neo component has been achieved. In the next
chapter the expression of the fusion gene in mammalian cells is discussed. Uncertainty as to whether the fusion gene functions as originally thought, means that one cannot be sure that this way of producing fusion genes can be applied universally. On the positive side, the possibility that insertion of the hprt sequence into the neo gene may have introduced a novel promoter leads to the suggestion that the intentional use of eukaryotic coding sequences that act as bacterial promoters could be used as an alternative way of constructing fusion genes.
CHAPTER 6

TARGETED INTEGRATION OF pAH(M⁻) INTO THE HPRT GENE
6.1. Introduction

The previous chapters have described the construction of the plasmids pAH and pAH(M'), and the characterisation of the cells and of the transfection system. The plasmids pAH and pAH(M') were constructed and tested in E. coli. The promoterless form of the fusion neo gene appeared to be expressed as effectively as with the promoter, and this was explained by an increase in the copy number of the plasmid. To use this plasmid to target integration into the hprt gene, it was important that it only confers G418 resistance when coupled to the Mmt promoter. In this chapter, I show that pAH expresses a functional fusion gene that confers G418 resistance when integrated into the genome of mammalian cells, and is expressed from the Mmt promoter.

The plasmid, pAH(M') was then used to target integration into the hprt gene. Initially this was attempted into the hprt gene expressed from p4aA8, when the plasmid was co-introduced into A9 cells along with various forms of pAH(M'). It was demonstrated that the two plasmids could undergo homologous recombination, and G418 could be used to select these events. Having demonstrated that the pAH(M') can be used to target integration, I describe my attempt at targeting the integration of pAH(M') into hprt genes contained within the chromosomal DNA.

6.2 Introduction of pAH into mammalian cells

The plasmid pAH was microinjected into A9 cells in one of the following combinations: pAH alone; pAH(M') alone; a 1:1 molar ratio of pAH(M') and p4aA8 or a 1:1 molar ratio of pREV27 plus p4aA8. The injected cells were selected either with G418 or HAT medium, and the results are shown in Table 6.1.
Table 6.1 Microinjection of pAH, pAH(M−) and pREV27 with or without p4aA8 into A9 cells

<table>
<thead>
<tr>
<th>Plasmid.</th>
<th>Selection</th>
<th>No. of colonies /cells injected</th>
<th>% of successful injections</th>
</tr>
</thead>
<tbody>
<tr>
<td>pAH</td>
<td>G418</td>
<td>7/320</td>
<td>2.2</td>
</tr>
<tr>
<td>pAH(M−)</td>
<td>G418</td>
<td>0/300</td>
<td>0.0</td>
</tr>
<tr>
<td>pAH(M−) + p4A8</td>
<td>G418</td>
<td>0/170</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>HAT</td>
<td>10/160</td>
<td>6.3</td>
</tr>
<tr>
<td>pREV27 + p4A8</td>
<td>G418</td>
<td>0/160</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>HAT</td>
<td>18/160</td>
<td>11.3</td>
</tr>
</tbody>
</table>
Figure 6.1  Analysis of the cell line M1:1 generated by microinjection of pAH into A9 cells

5 μg of DNA was digested with BamHI, BglII or BamHI and BglII, separated on a 0.7% agarose gel, blotted onto Biotrace-RP, hybridised to a $^{32}$P labelled neo-specific probe, and autoradiographed for 5 days.

Included on the gel were BamHI/BglII digests of pML2dMmtneo and pAH. These show that the BamHI/BglII neo-hybridising fragment of M1:1 has a similar size to that of pAH. The slight difference in mobility is most likely an artefact of electrophoresis and would be eliminated by mixing the plasmid DNAs with cellular DNA before electrophoresis.

BamHI digested pAH was included on the gel in amounts equivalent to 1, 5 and 10 copies of plasmid per cell.

The sizes shown are in Kb.
<table>
<thead>
<tr>
<th>pML2dMnt/neo</th>
<th>BamHI/BgIII</th>
<th>BgIII</th>
<th>pAH BgIII/Bam</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BamHI</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>1c</th>
<th>5c</th>
<th>10c</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>23.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>9.42</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.56</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.02</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

probe: -neo
These results demonstrated that the neo coding sequence of pAH was expressed in mammalian cells resulting in the appearance of G418-resistant colonies. Some of these colonies were picked and grown up. They were found not to be HAT-resistant, indicating that the hprt component of the fusion gene does not have significant hprt enzyme activity. In no case did injection of pAH(M'), either on its own or in mixtures with other plasmids, result in G418-resistant colonies. Coinjection of pAH(M') and p4aA8 resulted in HAT-resistant colonies, indicating that pAH(M') must also have been delivered to the nucleus. This demonstrated that the Mmt promoter was required for normal expression of the fusion gene in mammalian cells. Injection of pREV27, which contains the Mmt promoter but has the hprt sequence reversed, did not result in G418-resistant colonies. This suggested that the correct hprt reading frame is required for the production of a protein with the activity of neo, supporting the notion that G418 resistance results from translation of the entire fusion gene. The inability of pREV27 to express a functional neo protein also made it unlikely that the G418-resistant colonies produced by microinjection of pAH resulted from loss of the hprt region from the hprt/neo fusion gene.

DNA was purified from G418-resistant colonies isolated after injection with pAH and samples were digested with either BamHI, BglII or both enzymes. Each enzyme alone should cut the plasmid once, and together they should excise a 2.7 Kb fragment that contains the hprt/neo fusion gene. These digests were separated by electrophoresis, blotted and hybridised to a neo-specific probe. An autoradiograph of the resulting pattern of hybridisation for one of these clones (M1:1) is shown in Figure 6.1. BamHI and BglII alone each produce two bands. Both contain a 7.2 Kb band that corresponds to the size of linearised pAH. In addition they contain a unique band that hybridises less
Figure 6.2 Proposed structure of the integrated plasmid DNA of the cell line M1:1

This figure illustrates how three copies of pAH may have integrated as a head-to-tail concatamer into the genomic DNA of the cell line M1:1. Plasmid DNA is shown as a single line and genomic DNA is shown as dotted line, and also included are the positions of the BamHI and BglII sites and the Mmt, neo and hprt (filled areas) sequences.

Underneath the proposed map is illustrated the fragments that would hybridise to a neo-specific probe. The fragments marked (*) are unique junction fragments.
intensely. The simplest interpretation of these patterns is that three or more plasmid molecules have integrated in a head-to-tail concatamer to produce full length linears and unique junction fragments upon digestion with any "single-cut" enzyme. This is illustrated in Figure 6.2. All plasmids of the G418-resistant colonies investigated contained a BamHI-BglII neo-hybridising fragment that is approximately the same size as the equivalent BamHI-BglII fragment of pAH. These results indicated that following integration pAH contains an intact hprt/neo fusion gene that was expressed, conferring G418 resistance.

6.3 Homologous recombination between pAH(M-) and p4aA8 following microinjection

The design of these plasmids was such that when co-injected, homologous recombination would have been expected to result in a functional neo gene by linking the neo coding sequence of pAH(M-) to the SV40 promoter that controlled expression of hprt in p4aA8. In the experiment described above, co-injection of pAH(M-) and p4aA8 into A9 cells produced no G418-resistant colonies. The absence of G418-resistant colonies suggested that either homologous recombination between the two molecules occurred at a frequency too low to detect in these experiments, or that the products of the recombination did not produce a functional product.

In the experiment detailed in Table 6.1 both plasmids were in the supercoiled circular form, but from the results of others (see Section 1.8) linearisation within the region of homology would be expected to increase the frequency of recombination at this site. pAH(M-) was linearised within the hprt region by digestion with XhoI. An additional linear form was made by digesting pAH(M-) with XhoI and BamHI and isolating the resulting 2.6 Kb fragment which contained the
Table 6.2 Targeting the integration of the hpert/neo fusion gene into the hpert gene of p4aA8 after co-injection into A9 cell nuclei

<table>
<thead>
<tr>
<th>Input DNA</th>
<th>Selection</th>
<th>No. of colonies /injected cells</th>
<th>% of successful injections</th>
</tr>
</thead>
<tbody>
<tr>
<td>p4aA8</td>
<td>HAT</td>
<td>7/100</td>
<td>7.0</td>
</tr>
<tr>
<td>pAH(M⁻), uncut+ p4aA8</td>
<td>HAT</td>
<td>16/200</td>
<td>8.0</td>
</tr>
<tr>
<td></td>
<td>G418</td>
<td>0/200</td>
<td>0.0</td>
</tr>
<tr>
<td>pAH(M⁻), XhoI linear + p4aA8</td>
<td>HAT</td>
<td>5/52</td>
<td>9.6</td>
</tr>
<tr>
<td></td>
<td>G418</td>
<td>17/300</td>
<td>5.7</td>
</tr>
<tr>
<td>XhoI/BamHI pAH(M⁻) + p4aA8</td>
<td>HAT</td>
<td>6/100</td>
<td>6.0</td>
</tr>
<tr>
<td></td>
<td>G418</td>
<td>6/400</td>
<td>1.5</td>
</tr>
</tbody>
</table>
neo sequence linked to the major part of the hprt insert. This fragment differed from the linearised pAH(M⁻) in that it did not contain any plasmid sequences and only had homology to the hprt sequence of p4aA8. The circular and linear forms of pAH(M⁻) and the XhoI/BamHI fragment of pAH(M⁻) were co-injected with p4aA8. The results are shown in Table 6.2.

The number of HAT-resistant colonies resulting from injection of the plasmid p4aA8 is a measure of the efficiency of the microinjection process for each mix of DNA. Injections of mixtures of intact plasmid results in about 10% of the cells forming stable colonies. After linearising pAH(M⁻) within the region of homology to the hprt gene, G418-resistant colonies were formed at a frequency of about 60% of that for HAT resistance. Interestingly, the XhoI/BamHI fragment of pAH(M⁻) also gave G418-resistant colonies, but at a frequency of 25% of that for HAT resistance, which represents a reduction of just over twofold of that for linearised pAH(M⁻). As the results in Section 4.3 showed, a promoterless neo gene is only 1% as efficient at transforming A9 cells to G418 resistance as one with a promoter. Thus, a maximum of only 1 in 1000 of those cells injected with pAH(M⁻) would be expected to become G418-resistant as a consequence of integration into chromosomal DNA.

G418-resistant colonies, from the p4aA8/pAH(M⁻) coinjection experiments detailed in Table 6.2, were isolated, grown up and DNA made from them. These DNAs were digested by BamHI, separated on a 0.7% agarose gel and blotted for analysis by hybridisation. If homologous recombination had occurred between the hprt sequences common to both plasmids, a 2.9 Kb BamHI fragment which contained both the neo coding sequence and part of the SV40 promoter would have been produced (Figure 6.3). The blot was first hybridised with a probe specific to
Figure 6.3  The predicted products of homologous recombination between p4aA8 and pAH(M⁻)

On the left, this figure illustrates the product expected after a single reciprocal exchange between the hprt sequences (filled areas) of p4aA8 and pAH(M⁻). Included in the figure are the sizes of the BamHI fragments of p4aA8, pAH(M⁻) and the expected product.

On the right, this figure illustrates the expected patterns of hybridisation to either a probe specific for the SV40 promoter of p4aA8 or the neo fragment of pAH(M⁻) after digestion with BamHI. The unique feature of the product of homologous recombination is that it would be expected to contain a 2.9 Kb BamHI fragment that hybridises to both SV40 promoter and neo probes.
Figure 6.4  Analysis of the G418-resistant cell lines generated by co-injection of p4aA8 and pAH(M⁻)

10 μg of DNA from each cell line was digested with BamHI, separated on a 0.7% agarose gel, blotted, hybridised to either a probe specific for the SV40 promoter of p4aA8 (A) or the neo-specific probe (B), and autoradiographed for two weeks.

Cell lines M28, M51:1, M51:2 and M51:3 were established from G418-resistant colonies that formed after co-injection of circular p4aA8 and XhoI digested pAH(M⁻).

Cell lines M44 and M58 were established from G418-resistant colonies that formed after co-injection of circular p4aA8 and the XhoI/BamHI fragment of the hprt/neo fusion gene isolated from pAH(M⁻).

The 2.7, 2.9 and 3.0 Kb fragments are indicated (see Section 6.3).

The tracks marked M are Lambda HindIII fragments hybridised to ³²P-labelled Lambda DNA.
the region of the SV40 promoter expected to be contained in the p4aA8-derived region of the 2.9 Kb fragment, the probe was removed and then rehybridised with a neo-specific probe, in order to attempt to identify unambiguously the product of homologous recombination. To identify the SV40 region it was only possible to isolate as a probe a fragment that covered the entire SV40 promoter region. This meant that a second SV40-hybridising band, 3.0 Kb in size, representing the rest of p4aA8, would be expected. In addition it was expected that the SV40 promoter probe would be expected to hybridise to a 1.4 Kb BamHI fragment produced from hprt genes that had not integrated the plasmid pAH(M⁻).

The results of both hybridisations are shown in Figure 6.4. Cell lines M28, M51:1, M51:2, M51:3 were established from G418-resistant colonies that formed after co-injection of A9 cells with a mixture of linear pAH(M⁻) and p4aA8. In all four cases it is possible to identify the 3.0 Kb p4aA8-derived fragment, the variation in strength of the hybridisation being due to variation in copy number. In M51:2, a 2.9 Kb SV40 hybridising fragment was identified. When the membrane was rehybridised with the neo-specific probe, this 2.9 Kb fragment was also found to hybridise, fulfilling the prediction for the product of homologous recombination between hprt sequences of p4aA8 and pAH(M⁻). It may not have been possible to positively identify a similar fragment from the other DNAs since they had lower copy numbers of integrated plasmid and there was only a short region of homology between the probe and the SV40 sequence contained in the 2.9 Kb fragment. The existence of limits of detectability by hybridisation is indicated by the difficulty in detecting the 1.4 Kb fragment when it is present in low copy number (see the tracks for the cell lines M28 and M51:1). Thus, although it was not possible to identify a 2.9 Kb
fragment in cell lines M28, M51:1 and M51:3 with the SV40 probe, the fact that the neo-specific probe hybridised to a fragment of this size in all tracks suggests that homologous recombination may have generated a functional neo gene in all four cases.

In addition to the 3.0 and 1.4 Kb fragments the SV40 promoter probe hybridised to a 2.7 Kb fragment in all four cases. The presence of this fragment could be explained if a second homologous recombination event had occurred between the bacterial plasmid vector components of the plasmids. Both plasmids are based on pBR322, and therefore have homology in these regions as well as between the hprrt sequences, but the plasmid component of pAH(M) is pML2, which carries a 1370 bp deletion relative to pBR322. The SV40 promoter sequence of the p4aA8 plasmid and the deletion of pML2 act as markers flanking a 1.8 Kb region of homology within the plasmid vector sequences. A fragment of 2.7 Kb which hybridises the SV40 probe is indicative of the product of this recombination since it carries both of these sequence markers.

The BamHI digests of the DNAs of two of the cell lines (M28 and M51:3) contained many neo-hybridising fragments in addition to that of the 2.9 Kb fragment. The 5.4 Kb fragments in common with three of the cell lines (M28, M51:2, M51:3) could be unrecombined pAH(M'). The other fragments are best explained as rearrangements or deletions of the input plasmids produced by nonhomologous recombination, either between the injected plasmids as they concatamerise, or between plasmid and chromosomal DNA as the plasmids integrate into the genome. This latter possibility would suggest a large number of integration sites. It would be possible to establish the number of integration sites by using FIGE analysis as described in Section 3.4. After microinjection few rearranged bands are generally seen, so it may be that in these cases selection for those cells which were capable of
homologous recombination also selected for cells with high nonhomologous recombination activity.

The idea that each end of the linearised plasmid molecule may recombine independently is supported by the observation that co-injection of a XhoI/BamHI fragment of pAH(M⁻), which has homology at only one end, with p4aA8 was also able to produce G418-resistant colonies at a high frequency. The hybridisation pattern of DNAs from these cells were strikingly different (Figure 6.4: M44, M58) from those of the cells derived from microinjection of the linearised complete pAH(M⁻),(Figure 6.4: M28, M51:1, M51:2, M51:3). All of the BamHI fragments of lines M44 and M58 hybridising to the neo probe are of higher molecular weight than the 2.9 Kb fragment produced by homologous recombination of linearised pAH(M⁻) and p4aA8. For a BamHI fragment of 2.9 Kb to be present in these DNAs, the BamHI site with which the XhoI/BamHI fragment of pAH(M⁻) was originally excised would have had to have been reformed after integration into the target sequence. This did not appear to have occurred in either of the two lines investigated, so it must be concluded that the BamHI end of the XhoI/BamHI fragment of pAH(M⁻) did not recombine into a homologous site to restore the BamHI site. This was the case despite the presence of such a site in the target plasmid and would argue that the "sticky end" produced by BamHI cleavage does not contain a sufficient length of sequence to ensure homologous recombination. This end of the XhoI/BamHI fragment of pAH(M⁻) must have integrated through a process of nonhomologous recombination. A simple insertion of the XhoI/BamHI fragment of pAH(M⁻) into the hprt region can also be ruled out since such an event would be expected to generate a BamHI fragment of 3.4 Kb or less, and no fragments of this size were detected.
Is there any direct evidence that homologous recombination occurred between the hppt sequences of p4aA8 and the XhoI/BamHI fragment of pAH(M−)? The results of the hybridisations are difficult to interpret. It may be that some of the XhoI/BamHI fragments of pAH(M−) recombined by homologous recombination within the hppt gene and became linked to the SV40 promoter, but unfortunately the sensitivity of the SV40 probe was not high enough to detect a single copy fragment. The cell line M58 has a 8.0 Kb fragment which hybridises to both the SV40 and neomycin probes. This may be a candidate for the product of a homologous recombination, however the possibility that it was generated by nonhomologous recombination which linked a XhoI/BamHI fragment of pAH(M−) to a piece of DNA derived from p4aA8 that also contained sequences that hybridised the SV40 probe, cannot be excluded.

Another interesting feature of the hybridisation of the SV40 probe is the fragment of approximately 1.6 Kb. It is not clear from which of the original BamHI fragments of p4aA8 it was derived, or how it was created. The simplest explanation is that it represents a slightly larger version of the 1.4 Kb SV40 hybridising fragment seen in the cell lines M28, M51:1, M51:2 and M51:3, and is due to a chance insertion or minor rearrangement. The absence of the normal 1.4 Kb fragment would support this view.

6.4. Homologous recombination between pAH(M−) and p4aA8 following transfection using calcium phosphate precipitation

Co-transfections of A9 cells were carried out using circular p4aA8 together with either circular or XhoI linearised pAH(M−). The
Figure 6.5 Analysis of the G418-resistant cell lines generated by co-transfection of p4aA8 and pAH(M⁻)

10 ug of DNA from each cell line was digested with BamHI, separated on a 0.7% agarose gel, blotted, hybridised to either a probe specific for the SV40 promoter of p4aA8 (A) or the neo-specific probe (B), and autoradiographed for two weeks.

Cell lines HR12, HR13, HR14 and HR16 were established from G418-resistant colonies that formed after co-transfection of circular p4aA8 and XhoI digested pAH(M⁻).

Cell lines HR33, HR34, HR35, HR36 and HR37 were established from G418-resistant colonies that formed after co-injection of circular p4aA8 and circular pAH(M⁻).

The 2.7, 2.9 and 3.0 Kb fragments are indicated (see Section 6.4).

The tracks marked M are Lambda HindIII fragments hybridised to $^{32}$P-labelled Lambda DNA.
Probe: SV40 promoter

Probe: neo
following transfection frequencies were obtained when transfectants were selected for G418 resistance:

\[
\begin{align*}
\text{pAH}(M^-), \text{linearised, plus p4aA8:} & \quad 1.95 \times 10^{-4} \\
\text{pAH}(M^-), \text{circular, plus p4aA8:} & \quad 1.10 \times 10^{-4}
\end{align*}
\]

G418-resistant colonies were produced at the frequencies characteristic of those observed when a neo gene with an attached promoter is transfected. Circular plasmids also produced G418 colonies at a high frequency after transfection by calcium phosphate precipitation. Thus transfection by calcium phosphate precipitation appears to increase the frequency of homologous recombination between two covalently closed circular DNAs.

Four colonies from the transfection with linearised pAH(M-) and five colonies from the transfection with circular pAH(r') were picked at random and grown up. DNAs from these cells were analysed in the same way as those from lines generated after microinjection (Figure 6.5).

Transfection with the linearised plasmid produced similar results to those by microinjection with the linearised plasmid. Again, the 3.0 and 2.7 Kb SV40-hybridising fragments and the 2.9 Kb neo-hybridising fragments were present. The results differed in that there were more rearranged neo-hybridising fragments after calcium phosphate mediated transfection. The many different intensities of the bands indicate that replication/amplification may have occurred after or during the formation of the rearranged forms. As discussed in Chapter 3, this was a feature of the transfection system. Unfortunately, the SV40 probe gave only a weak signal, and so it was not possible to identify
positively the predicted 2.9 Kb SV40-containing fragment, or any rearrangements of p4A8.

DNA from lines transfected with the two plasmids in the circular form gave a significantly different hybridisation pattern. Overall there were fewer copies of the neo gene. In two cases (HR34 and HR36), characteristic 3.0 and 2.7 Kb SV40-hybridising fragments and the 2.9 Kb neo-hybridising fragments were present. This suggested that in the founder cells of these lines a similar process of recombination occurred resulting in G418 resistance. In the other three lines, there was no common pattern in the sizes of the neo-hybridising bands and so G418 resistance in these lines does not appear to have resulted from the same recombinational processes. It was possible however, that a homologous recombination initially brought the SV40 promoter and the neo fusion gene together, and then rearrangements changed the positions of the BamHI sites. It was not possible to tell if any of these neo fragments also contained the SV40 promoter, due to the lack of sensitivity of the probe on the blots of these DNAs. There was a faintly SV40-hybridising 2.7 Kb fragment in line HR35. As in all other cases that this was found a 2.9 Kb fragment had also been formed, this may be a case where a second nonhomologous recombination event had obscured the formation of a 2.9 Kb neo-hybridising fragment. The alternative explanation of G418 resistance in these three cell lines was that the neo gene became expressed as a result of nonhomologous recombination fortuitously placing the SV40 promoter into the correct alignment.

In general, the same processes of recombination appear to occur in most lines after transfection. The only major difference between the recombination after transfection as opposed to that occurring after microinjection is that circular plasmids recombine at a higher
frequency in the former case. There are two possible explanations for this:

i) After transfection, more DNA was introduced into the cell than by microinjection, and hence the frequency the homologous recombination increased, as described by Folger et al (1982).

ii) The calcium phosphate precipitation method is known to result in damage to DNA. It may be that during the course of transfection the circular molecules became linearised.

The results of the hybridisation analysis show that only a few copies of the neo gene became integrated, however it is not possible to infer from this that only a few copies of the plasmid entered the nucleus and so the first explanation cannot be discounted. The second explanation suggests that transfection delivers linearised plasmids to the nucleus. In contrast to the precise cut introduced by digestion with XhoI, during transfection by calcium phosphate precipitation, circular plasmids would be broken at random. The break in the hprt region may have acted to direct recombination to sequences around this point. In the case of the initially circular forms, random linearisation would result in only a proportion of the molecules being broken near enough to the hprt sequence for homologous recombination to occur through the hprt gene. Molecules broken in non-hprt regions may have preferentially recombined within the plasmid sequences or through processes of nonhomologous recombination.
Table 6.3 Microinjection of pAH and pAH(M⁻) into the cell lines A9, M13 and A2-4

<table>
<thead>
<tr>
<th>Cell line</th>
<th>Plasmid</th>
<th>Selection</th>
<th>No. of colonies /cells injected.</th>
</tr>
</thead>
<tbody>
<tr>
<td>A9</td>
<td>pAH</td>
<td>G418</td>
<td>7/320</td>
</tr>
<tr>
<td></td>
<td>pAH(M⁻)</td>
<td>G418</td>
<td>0/300</td>
</tr>
<tr>
<td>M13</td>
<td>pAH</td>
<td>G418</td>
<td>20/280</td>
</tr>
<tr>
<td></td>
<td>pAH(M⁻)</td>
<td>G418 + 6TG</td>
<td>0/5000</td>
</tr>
<tr>
<td>A2-4</td>
<td>pAH</td>
<td>G418</td>
<td>19/300</td>
</tr>
<tr>
<td></td>
<td>pAH(M⁻)</td>
<td>G418 + 6TG</td>
<td>0/5000</td>
</tr>
</tbody>
</table>
6.5. Homologous recombination into a chromosomal site

In the previous two sections, it has been established that selection on G418 can be used to isolate colonies, in which the integration of the plasmid pAH(M⁻) has been targeted to an hprt gene. In these examples the hprt gene was carried on a plasmid that was co-introduced into the nucleus with pAH(M⁻). The next question to ask was whether pAH(M⁻) could be used to target integration into an hprt gene contained within the chromosomal DNA. A targeted integration should, in addition to expressing the fusion gene from the hprt promoter, disrupt the hprt coding sequence and so generate a 6TG-resistant colony (see Figure 1.5).

The cell lines M13 and A2-4 were used for this experiment. M13 contains a single integrated copy of the plasmid p4aA8 (see Section 3.6), and therefore carries 480 bp of homology to pAH(M⁻) within its chromosome. The A2-4 cell line contains a single genomic copy of the human hprt gene, and therefore contains the same hprt coding sequences as the M13 line, but in A2-4 the sequences are interspersed between large introns, so that the length of homology at any particular potential integration site is reduced.

On the basis of the frequencies reported by Thomas et al (1986), it was felt that it might be possible to microinject sufficient cells to obtain colonies derived from cells that had targeted the integration of pAH(M⁻) into the chromosomal hprt gene. Table 6.3 shows the results of microinjection of XhoI linearised pAH(M⁻) into M13 and A2-4. Although injection of pAH produced G418-resistant colonies at 7.1% and 6.3% of the M13 and A2-4 cells injected respectively, no 6TG-G418-resistant colonies were obtained after injecting 5000 cells of M13 and A2-4 each. The subsequent report by Thomas and Capecchi (1987) would suggest that the frequency of targeted integration is at least
Table 6.4 Transfection of pAH(M⁻) into A9, M13 and A2-4 cells

<table>
<thead>
<tr>
<th>Cell line</th>
<th>Selection</th>
<th>No. of colonies/ x10⁷ cells transfected</th>
</tr>
</thead>
<tbody>
<tr>
<td>A9</td>
<td>G418</td>
<td>20</td>
</tr>
<tr>
<td>M13</td>
<td>G418</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>G418 + 6TG</td>
<td>2</td>
</tr>
<tr>
<td>A2-4</td>
<td>G418</td>
<td>107</td>
</tr>
<tr>
<td></td>
<td>G418 + 6TG</td>
<td>1</td>
</tr>
</tbody>
</table>
an order of magnitude lower than that originally expected. This may explain the inability to produce a colony resulting from targeted integration by microinjection.

$10^7$ M13 or A2-4 cells were transfected with XhoI linearised pAH(M'). The cells were selected either on G418 or on G418 and 6TG. The results are shown in Table 6.4. Transfection of the M13 cell line with pAH(M') resulted in two colonies that were both 6TG- and G418-resistant. This gave a potential frequency of a targeted integration of $2 \times 10^{-7}$, whilst the background of transfected cells that were only G418-resistant was an order of magnitude higher. This was consistent with the results expected from the transfection of the promoterless neo gene described in Chapter 4. These were picked and grown up to establish the cell lines HR29 and HR32.

Only a single 6TG- and G418-resistant colony was obtained as a result of transfecting pAH(M') into A2-4 cells. This gave a potential frequency of a targeted integration of $1 \times 10^{-7}$, in this case the background frequency of transfected cells that were only G418-resistant was two orders of magnitude higher. As shown in Section 4.3, there appears to be little difference between the frequency of G418-resistant cells after transfection with the promoterless gene in the plasmid p(Bam)neo of either A9 or A2-4 cells, and so the high background frequency observed with pAH(M') may simply be an anomaly of this control. If, however, the control did represent the true figure for all transfections of XhoI linearised pAH(M') into A2-4 cells, then the background frequency of 6TG- and G418-resistant colonies would be expected to be about $1 \times 10^{-7}$. This figure is high enough to cast doubt on whether this colony may have resulted from a targeted integration or is due to simultaneous loss of the hprt gene.
and expression from the promoterless hprt/neo fusion gene. This colony was picked and grown up to establish the cell line HR17.

Attempts to analyse the structure of the integrated plasmid in these cells by electrophoresis, blotting and hybridisation have failed to establish whether pAH(M') integrated by homologous recombination into the chromosomally located hprt gene. In the case of HR32 a neo-hybridising fragment which could have been the predicted 2.9 Kb fragment was detected. Since no hybridisation was achieved with the SV40 probe, it is not possible to conclude whether this 2.9 Kb neo-containing fragment had become linked to the SV40 promoter by homologous recombination. Further analysis would be necessary to clarify the situation.

6.6. Discussion

In this chapter, it has been demonstrated that pAH produces G418-resistant colonies when it is integrated into the chromosome of mammalian cells. The results obtained following microinjection are consistent with transcription occurring from the Mmt promoter and translation of the entire hprt/neo fusion gene. It therefore proved possible to design by computer a fusion gene for use in mammalian cells, and to construct and test it using E. coli as a model system.

By co-introducing the promoterless hprt/neo fusion gene contained in pAH(M') and a functional hprt gene contained in p4aA8, into A9 cells, it was shown that it was possible to target the integration of the hprt/neo fusion gene, that contains a 480 bp region of homology, into the hprt gene contained in a co-injected plasmid. Analysis of DNA from the resulting G418-resistant colonies identified a 2.9 Kb BamHI fragment that hybridised probes specific to the neo gene and the SV40 promoter region. This fragment was predicted as an outcome of
homologous recombination between the hpkt sequences in both plasmids. As established in previous work, cutting pAH(M) in the hpkt sequence increased the frequency of homologous recombination to a detectable level (see Section 1.8).

It was found that when the plasmids were introduced by calcium phosphate precipitation, G418-resistant colonies were produced at a high frequency when both plasmids were in the circular form. This could be explained if damage to the DNA, as a consequence of transfection, resulted in double strand breaks within the plasmids, thereby increasing the frequency of homologous recombination. Double strand breaks would be expected to be produced at random locations in either or both plasmids, so not all plasmids would be broken within, or near, the homologous hpkt sequences. Loose ends created by random breaks within the vector portion of the plasmids may also undergo homologous recombination, but these would not form a functional neo gene and so would not be selected for. In addition many junctions would be created through processes of nonhomologous recombination, some of which may link the SV40 promoter to neo coding sequence. This is supported by the observation that only two of the five colonies picked from a co-transfection with circular forms of both p4aA8 and pAH(M) produced hybridisation patterns consistent the generation of a 2.9 Kb BamHI neo hybridising fragment by homologous recombination, and so the other three must have resulted from nonhomologous recombinations. This is consistent with the observation that after transfection of pML2dMMtneo using the calcium phosphate precipitation method, many rearranged plasmid fragments are found, (Figure 3.5b).

It was also found that the XhoI/BamHI fragment of pAH(M), which contains homology to the hpkt gene only at a single end and no homology to the plasmid sequences of p4aA8, also gave G418-resistant
Figure 6.6  Two pathways which may produce a plasmid with the 2.7 Kb SV40 promoter-hybridising fragment

A: A single reciprocal exchange between the hprt sequences (filled areas) would produce a heterodimer. A second reciprocal exchange between the plasmid vector sequences (single line) of the heterodimer would reform two plasmids, one of which contains the 2.9 Kb neo-hybridising fragment and the 2.7 Kb SV40 promoter-hybridising fragment.

B: Two independent reciprocal exchanges may occur simultaneously, one between the hprt sequences and one between the plasmid vector sequences. This would produce a plasmid that contains the 2.9 Kb neo-hybridising fragment and the 2.7 Kb SV40 promoter-hybridising fragment in a single step. It may also produce a reciprocal product of unpredicted structure.
colonies after co-injection with p4aA8. In these colonies, the neo coding sequence was also considered to have become expressed as a result of homologous recombination between the hpert sequences.

After the identification of the predicted 2.9 Kb BamHI fragment, the most interesting feature of the homologous recombination events was that a novel 2.7 Kb BamHI fragment, which hybridises to the SV40 promoter probe, was formed. This was only found after p4aA8 had been introduced with the complete promoterless hpert/neo fusion gene plasmid, and, in all but one case, is associated with the generation of the 2.9 Kb fragment expected for recombination between hpert sequences (see Figures 6.4, 6.5). It was concluded that the 2.7 Kb fragment was produced by a second reciprocal exchange event occurring between the homologous regions of the bacterial vector components of each plasmid. The simplest way to consider the interaction between the two plasmids is to imagine that only one molecule of each plasmid is present in the nucleus of a cell. The two recombination events could be imagined to proceed in either of two ways. Homologous recombination between the hpert sequences could generate a structure in which one copy of each plasmid is joined together to form a heterodimer and then a second homologous recombination event would produce a novel structure which contains the 2.7 Kb BamHI fragment in addition to the functional hpert/neo fusion gene (Figure 6.6a) The alternative explanation is that two homologous recombination events may proceed simultaneously, one between the regions of hpert sequence homology and one between the regions of plasmid vector sequence homology, to produce a novel plasmid which contains the 2.7 Kb BamHI fragment in addition to the functional hpert/neo fusion gene (Figure 6.6b). In fact more than one of copy of each plasmid may undergo recombination, and this is expected to generate a concatamer of plasmids, the underlying
interactions, however, are adequately represented in a situation where only two plasmids interact. It is possible that in the first explanation the second recombination event may occur after the heterodimer (or concatamer) has integrated. It has been shown by Folger et al. (1982) that plasmids concatamerise rapidly after injection into the nucleus and that after integration the structure of these concatamers is highly stable. This would argue that the recombination events described in Sections 6.3 and 6.4 of this chapter occurred between molecules whilst in the extrachromosomal state soon after entry into the nucleus.

Most investigations have concentrated on the changes which occur within the immediate area of interaction of the two DNA molecules during homologous recombination, and so do not give information concerning the fate of the flanking vector sequences. In two cases, the isolation of heterodimers, which contained both products of a homologous recombination event, have been reported (Ayares et al., 1986; Seidman, 1987). Heterodimers, however, were uncommon and the majority of the plasmids isolated in these experiments only contained one of the predicted possible products of homologous recombination. There are also reports that intramolecular homologous recombination can generate monomer units from oligomers of SV40 (Wake & Wilson, 1980), release wildtype SV40 DNA from SV40/plasmid heterodimers (Subramani & Berg, 1983), and generate a single plasmid containing a wildtype large T antigen gene from a heterodimer which possesses two large T antigen genes with different mutations (Chakrabarti & Seidman, 1986). These observations suggest that a heterodimeric intermediate may be formed by homologous recombination between two plasmids and that homologous recombination may resolve the heterodimeric structures into their monomeric components. These observations, however, do not
indicate whether during the process of recombination, all plasmid recombination proceeds through an obligatory heterodimeric intermediate which is then rapidly converted to monomers, or whether heterodimers are formed by a low frequency mechanism and novel monomer plasmids by a different high frequency mechanism.

The alternative explanation to an obligatory heterodimeric intermediate is that homologous recombination occurred simultaneously between the hprt sequences and between the plasmid vector sequences. If this was the case it would suggest that both regions of homology may recombine independently of one another. This is consistent with the observation that the XhoI/BamHI fragment of pAh(M') may have undergone homologous recombination with p4aA8, despite only having sequence homology at the 5' end.

A double strand break may stimulate recombination in the region within which it is situated, in two ways, either the ends are directly involved in the recombination or recombination is enhanced in the regions that surround the break. If each end of XhoI linearised pAh(M') recombines separately, and recombination may only occur at the ends of the molecule, in order to generate both the 2.9 Kb and the 2.7 Kb fragment then the ends of the linear molecule must be degraded and hence the plasmid vector sequences at the 5' side of the XhoI cut exposed. This degradation would have to remove at least 215 bp. of sequence 5' of the XhoI site. Homologous recombination has been shown to proceed despite the presence of heterologous sequences (Brenner et al, 1986).

If it is the case that the ends of linearised pAh(M') had each recombined independently with p4aA8, there are still two possible products. In both cases, the homologous hprt sequence that is 3' of
the XhoI site would have recombined to form the 2.9 Kb BamHI fragment. The regions of pAH(M-) DNA that are 5' of the XhoI site may have recombined within the sequences that are either homologous to the hprt gene or homologous to the plasmid vector. If recombination had occurred between the vector sequences, then as stated above, one end of pAH(M-) would have been linked to the hprt sequence and the other end would have been linked to the plasmid vector sequence so a novel plasmid that contained the 2.7 Kb fragment would be formed. If, however, both ends had recombined between the hprt sequences, then a heterodimer would have been formed. This could be through both ends interacting at a single site by a mechanism such as that proposed in the double strand break model (see Figure 1.3). Alternatively, both ends could interact independently and so, for example, two separate Holliday junctions (see Figure 1.2) may form at the ends which, if they resolve in different planes (as in the double strand break model), a single reciprocal exchange would create a linear heterodimer. This then may be recircularised by ligation. It is possible therefore that, in fact, recombination may proceed by both of the pathways described in Figure 6.6, depending on exactly which regions of the two plasmids are involved in homologous recombination. The advantage of this argument is that each end of a piece of DNA may recombine by different mechanisms, so in the case of the XhoI/BamHI fragment of pAH(M-), homologous recombination may have occurred within the 5' hprt sequences and nonhomologous recombination occurred at the 3' end of the neo gene. It must be pointed out that recombination by a process of single strand annealing (see Figure 1.4) may also create the structures described above, however a second double strand break must form at a high frequency in the hprt sequences of p4aA8.
When considering various pathways of homologous recombination between two free plasmids, the final products of recombination may be the same irrespective of whether the pathway proceeds via a heterodimeric intermediate or one double recombination. The outcomes of the two pathways described in Figure 6.6, however, will have different consequences when the integration of a plasmid is targeted into a chromosomal gene. If both ends of the integrating DNA are constrained to recombine at a single point (the equivalent of having an obligatory heterodimer intermediate between two plasmids undergoing homologous recombination) then integration can only occur by insertion of the plasmid into the chromosome. If, however, the regions at the ends of the integrating plasmid DNA can recombine independently and the plasmid has sequences which are homologous to non-adjacent sites in the target DNA, then the process of integration will replace the region of the target that is contained between the two sites. This latter mechanism provides the opportunity to replace one region of the chromosomal DNA with another piece of DNA, and has been observed to occur with a vector designed by Thomas and Cappecchi (1987). The "replacement" vector was designed to replace a region of exon 8 in the genomic hprt gene with a neo gene. In it, a neo gene was flanked on the 5' side by DNA which had homology to exons 6 and 7 of the mouse hprt gene and on 3' side by DNA with homology to exon 9 of the mouse hprt gene. This was observed to replace exon 8 with neo in every (23/23) case analysed. They also designed an insertion vector which had the neo gene flanked by the same DNA, but in this experiment the 5' and 3' ends of the hprt regions were joined and the molecule cut within exon 7. This was expected to recombine at a single site, insert the neo gene and generate a duplication of the chromosomal DNA. In the majority of cases this proceeded as expected, however in 3/12 cases
analysed, this vector unexpectedly integrated in the same manner as the replacement vector. In these latter cases one homologous recombination may have occurred at the break and the other within the region of homology to exon 9. These exceptions support the notion that two regions of the integrating DNA may recombine independently.

I have attempted to use linearised pAH(M⁻) to target integration into an hprt gene situated within the chromosomal DNA. Two different cell lines, M13 and A2-4 were used, differing from each other in that M13 had the greater length of homology between the plasmid and chromosomal sequence, and that M13 had been generated by microinjection of p4aA8 into A9 cells. It may have had therefore an hprt gene situated in a region of the genome which was more accessible to integrating DNA. No colonies that had potential targeted integrations of pAH(M⁻) were isolated after microinjection into either cell lines. This suggested that targeted integration of pAH(M⁻) by microinjection occurred at a frequency too low to detect.

Transfection by calcium phosphate precipitation gave two colonies from the M13 line and one colony from the A2-4 line which were 6TG- and G418-resistant, suggesting that they may have undergone targeted integration. These arose at a frequency of about 1x10⁻⁷, which is about 1/1000 of the frequency of cells expected to integrate DNA. This figure is comparable with those observed by Smithies et al, (1986) and the optimum frequencies obtained by Thomas & Capecchi, (1987). It is lower than that obtained by Doetschman et al, (1987) unless the ES cell line used can potentially integrate DNA at a frequency of about 1x10⁻³ after electroporation. As the length of homology between the integrating DNA and the target is at least greater than 2.5 Kb in these reports, it would be of interest to obtain definite proof that pAH(M⁻), which has only 480 bp, is also
able to give targetted integration at this frequency. In targetted integration of pAH(M) into the M13 line it would also be interesting to see whether integration occurs through a process of insertion into a single site within the hprt sequence or by replacement of the 3' region of hprt gene and plasmid sequences with a region of pAH(M).
CHAPTER 7

DISCUSSION
Discussion

In this project it has been possible to construct a fusion gene where the coding sequence of human hprt cDNA is inserted into the 5' region of the neo coding region. The Wisconsin (UWGCG) sequence analysis software package was used successfully to model the effect on translation of insertions into coding sequence of the neo gene. This enabled the construction of pLink and then selection of a fragment that would correct the frame-shift mutant that pLink contained. By modelling in the computer and then introducing KpnI sites at suitable sites by oligonucleotide site directed mutagenesis, as done for the 5' site of the hprt insert, it should be possible to construct other fusion genes in the same manner. These can be rapidly tested in E. coli to see if they have neo activity. In addition to inserting a pre-selected fragment into pLink, it would also be possible to "shot-gun" clone fragments into it, and use kanamycin selection to isolate functional fusion genes in E. coli. By use of the neo gene in conjunction with the Mmt promoter it has proven possible to use a single plasmid to make a fusion gene that can be expressed in both prokaryotes and eukaryotes. This allowed construction of the fusion gene for a mammalian cell system to be done in E. coli, and hence reduced the time taken to test whether the construct functioned.

The advantage of using the neo gene in the fusion is that cells which express it can be selected on medium containing G418. It is possible therefore to express a single gene which has the activity of the neo gene and also contains a polypeptide derived from a second protein. This may be useful in situations where it
is necessary to directly select for the presence of a fusion protein. For example, it would be possible to create a combined fusion protein and selectable marker which could be used in virus vectors with narrow packaging limits, such as SV40 or retroviruses. As described by Reiss et al (1984b), the activity of the neo gene can be assayed by its ability to phosphorylate kanamycin using gamma-$^{32}$P-ATP as a substrate, and therefore used as an indicator of gene expression, in the same way as the chloramphenicol acetyl transferase gene. It should be possible to use pLink to investigate the effects on the neo activity of inserting extra polypeptide residues into the neo gene in both prokaryotes and eukaryotes. In conclusion, although more work would be required to investigate how pAH functions in E. coli and mammalian cells, this type of fusion gene has potential uses in areas other than as a vector for the selection of targeted integration.

It has been shown that the promoterless form of the hprt/neo fusion gene can be used to target integration into an hprt gene carried on a co-introduced plasmid. These results show that this process occurs at a high frequency when the plasmid is cut within the homologous sequence, which is consistent with observations made by other workers. The analysis of the products of this form of homologous recombination suggested that both ends of a linear molecule recombine independently. This is consistent with the observation that a DNA fragment with homology to the target gene only at one end can still recombine at a high frequency, and with the possible mode of action of the "replacement" vector of Thomas and Cappechi (1987). This situation can be analysed further to give more information concerning the events occurring during
recombination. For example further digests may reveal whether the double strand break model or the single strand annealing model is responsible for the products observed. It may also be possible to make further constructs, such as a heterodimer of p4aA8 and pAH(M⁻), observe their fate after microinjection and so test possible intermediate structures formed in the recombination pathway.

It has not been possible to demonstrate conclusively targeted integration into an hprt gene situated within the chromosome. It, however, would be of interest to confirm whether this is possible as this would demonstrate the use of shorter lengths of homology, such as between cDNAs and their genomic counterparts, than those used at present for targeted integration. The function of many genes, such as the mouse genes isolated by virtue of cross-homology to the "finger"-structures of Krüppel (Chowdhury et al, 1987), is unknown. The use of targeted integration to disrupt these genes would be a useful method of establishing the function of these genes. The ability to target integration into a gene using only its cDNA, or even a cDNA library would increase the effectiveness of this technique.

The experiment attempted in this project could be improved by the use of a cell line that does not lose the target hprt gene at frequency as high as that in the A2–4 cell line. This would reduce the background of 6TG–G418-resistant cells and so aid in the isolation of those cells that have undergone targeted integration. The transfection system could also be improved. For example, electroporation, as used by Smithies et al (1985),
Thomas and Capecchi (1987) and Doetschman et al. (1987), would be expected to be a more consistent method for the introduction of DNA into the cell and therefore to increase the overall transfection frequency. It would also allow targeted integration to be attempted in cell lines not amenable to transfection by calcium phosphate precipitation.

If the targeted integration of pAH(M) could be achieved at a frequency as high as 1 in 1000 cells which integrate DNA, it may be possible to make a plasmid of the same design in order to target integration into genes where it is not possible to select for their loss. From the transfection frequencies of the promoterless neo gene established in Chapter 4, 1 in 10 colonies analysed would be expected to result from a targeted integration, if selection was not used for loss of the target gene. The other 90% of the colonies would result from expression of the neo gene through nonhomologous recombination processes. This would enrich the number of targeted integrations to a level where colonies could be picked at random and screened by electrophoretic analysis.

In this project, and the reports in Section 1.9, the cells undergoing targeted integration have been selected from a constant background of colonies undergoing nonhomologous recombination, and the vectors have been designed accordingly. The other approach is to increase the ratio of homologous to nonhomologous recombination. In one report, it has been shown that the frequency of homologous recombination between linearised plasmids is ten to fifteenfold higher in early to mid S-phase than in other phases of the cell cycle (Wong & Capecchi, 1986).
As the frequency of nonhomologous recombination is constant throughout the cell cycle, the ratio of homologous to recombination is at its highest at this point. By synchronising the cell cycle of the target cells and, at the appropriate time, introducing a vector of a design similar to the one used in this project, it may be possible to increase the proportion of colonies that have undergone targeted integration.
REFERENCES
REFERENCES


BIRNBOIM, H.C. & DOLY, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. NAR 7 1513.


151


